REPORT DOCUMENTATION PAGE	2	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
27-08-2014	Final Report	26-Sep-2012 - 25-Feb-2014

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Final Report: Recognition of In-Vehicle Group Activities (iVGA): Phase-I, Feasibility Study	W911NF-12-1-0591
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS Amir Shirkhodaie	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Tennessee State University 3500 John A. Merritt Blvd Nashville, TN 37209 -1561	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO 11. SPONSOR/MONITOR'S REPORT NUMBER(S) 63085-MA-H.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

The objective of this proposed research effort is to conduct a feasibility study of in-vehicle Group Activities (iVGA) and develop suitable context-based taxonomy and ontology schemas for coherent analysis and inferencing of such activities. The prime objective of feasibility study is to identify suitable automated computer vision-based techniques enabling low false alarm rate identification of iVGA. Another objective is to develop a suitable virtual simulation environment as a test bed for staging appropriate context-based scenarios exhibiting different iVGA for testing and evaluation of newly developed ontology, algorithms and techniques. The goal is map visual

15. SUBJECT TERMS

In-Vehicle Group Activity, Taxonomy, Ontology, Image Processing, Imagery Data Inferencing and Annotation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Amir Shirkhodaie
UU	UU	UU	UU		19b. TELEPHONE NUMBER 615-963-5396

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report: Recognition of In-Vehicle Group Activities (iVGA): Phase-I, Feasibility Study

ABSTRACT

The objective of this proposed research effort is to conduct a feasibility study of in-vehicle Group Activities (iVGA) and develop suitable context-based taxonomy and ontology schemas for coherent analysis and inferencing of such activities. The prime objective of feasibility study is to identify suitable automated computer vision-based techniques enabling low false alarm rate identification of iVGA. Another objective is to develop a suitable virtual simulation environment as a test bed for staging appropriate context-based scenarios exhibiting different iVGA for testing and evaluation of newly developed ontology, algorithms and techniques. The goal is map visual observations of iVGA to annotated messages describing spatiotemporal behaviors exhibited in the confined space of a vehicle in lieu of physical obstructions and inevitable occlusions. In this project, partially observable spatiotemporal imagery data are analyzed to reason, predict, and discriminate group activities occurring in confined spaces.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

<u>Received</u> <u>Paper</u>

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

<u>Received</u> <u>Paper</u>

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received	Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT SUPPORTED	Discipline
New Entry•Azin Poshtkar	0.50	
FTE Equivalent:	**0.50**	
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT SUPPORTED	National Academy Member
Amir Shirkhodaie	0.05	
FTE Equivalent:	**0.05**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT SUPPORTED	Discipline
David A. Potter Jr.	0.25	Mechanical Engineering
FTE Equivalent:	**0.25**	
Total Number:	**1**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 1.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 1.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 1.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 1.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 1.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

This report presents our initial feasibility study which presents a systematic approach for analysis of target entities performing activities inside a vehicle, while being observed by a remote surveillance camera. This effort established appropriate and novel ontology for state-space representations of target entities' postural configurations inside a vehicle as well as their physical limitation – particularly, in terms of their kinematic constraints that may result impossible/unrealistic postural configurations of arms during normal activities/operations. To better visualize the admissible and impermissible postural configurations, a virtual model of a humanoid was developed using IRIS software (IRIS is developed by the PI outside of scope of this project). The humanoid model is a 32-degree-of-freedom robotic model with kinematic motions similar to those of humans. This model was utilized to simulate different human postural configurations in the virtual environment. The developed ontology were used for verification of admissible and inadmissible postural configurations. Furthermore, suitable image processing (IP) techniques were developed for automatic background segmentation, hands, arms, and head detection and tracking. Several novel image processing techniques were explored for skin detection and segmentation.

Technology Transfer

In this project, we held 13 monthly teleconference meetings with the ARL technical monitors at ARL. This frequent interaction with ARL has been significantly instrumental to transfer technology. This cooperative technology transfer effort will be maintained to assure the successful achievement of this mission. Furthermore, technical papers is currently being prepared by publication in the SPIE 2015 security and defense conference, Baltimore, MD.

FINAL PROJECT REPORT

Project Title:
Recognition of In-Vehicle Group Activities (*iVGA*) - Phase–I, Feasibility Study

ARO Program Manager:
Dr. Liyi Dai, Information Processing and Fusion, US ARO

Proposal Number:
63085MAH

Agreement Number:
W911NF-12-1-0591

Project Period:
9/25/2012 – 2/25/2014

Principal Investigator:
Dr. Amir Shirkhodaie
Department of Mechanical and Manufacturing Engineering
Tennessee State University
3500 John A. Merritt Blvd.,
Nashville, TN 37209
Tel: 615-963-5396
Email: ashirkhodaie@tnstate.edu

Date:
February 15, 2014

Table of Contents

Topic	Title	Page
1.	Acknowledgement	3
2.	Project Scope	4
3.	Introduction	4
4.	Project Motivation	4
5.	Scope of this Feasibility Study	5
6.	Research Accomplishments	5
	6.1 Development of Space-state Model for Occupants Inside a Vehicle	5
	6.2 Development of Ontology for Inferring Occupants' Postural Configurations Inside Vehicle	7
	6.3 Development of Arms Kinematic Configuration Admissibility (AKCA) Decision Trees	20
	6.4 Development of New Algorithms for Vehicle Windows Detection	24
	6.5 Development of New Algorithms for Human Skin Detection	26
	6.6 Development of A New Algorithm for Isolation of Human Body Part From Dynamic Videos	28
	6.7 Development of A New Algorithm for Tracking Human Body Parts Tracking	29
	6.8 Development of A New Algorithm for Initial Detection of Human Body Parts Based on SURF Image Processing Detection	35
	6.9 Development of Another Robust Algorithm for Hand Color Skin Detection Based on RBG Mean-Shift Technique	36
	6.10 Development of Two Competing Techniques for Hand Gesture Classification	39
	6.11 Development of a Frame Work for Enhancement of This Feasibility Study Toward Phase I of this Project	44
7.	Scientific Barriers	45
8.	Scientific Significance	45
9.	Scientific Accomplishments	46
10.	Collaboration and Leveraged Funding	46
11.	Technology Transfer	46
12.	Future Research Plan	46
13.	Anticipated Scientific Accomplishments	47
14.	Students Involvements	47
15.	Conclusion	47
16.	References	48

Acknowledgements

This research project was supervised by two technical monitors, Dr. Alex Chan, Dr. Shaun Hu from US Army Research Laboratory. The project contained a number of monthly teleconference call for exchange of technical details of this project. The valuable discussion with both technical monitors are acknowledged – Particularly, Dr. Alex Chan for his initial consideration and funding allocation on this project.

Furthermore, the support of US Army Research Office (ARO), Project Manager, Dr. Liyi Dai, from Information Processing and Fusion division is acknowledged.

2. Project Scope

The objective of this proposed research effort is to conduct a feasibility study of *in-vehicle Group Activities (iVGA)* and develop suitable context-based taxonomy and ontology schemas for coherent analysis and inferencing of such activities. The prime objective of feasibility study is to identify suitable automated computer vision-based techniques enabling low false alarm rate identification of iVGA. Another objective is to develop a suitable virtual simulation environment as a test bed for staging appropriate context-based scenarios exhibiting different iVGA for testing and evaluation of newly developed ontology, algorithms and techniques. The goal is map visual observations of iVGA to annotated messages describing spatiotemporal behaviors exhibited in the confined space of a vehicle in lieu of physical obstructions and inevitable occlusions. In this project, partially observable spatiotemporal imagery data are analyzed to reason, predict, and discriminate group activities occurring in confined spaces.

3. Introduction

The asymmetric adversarial group activities are generally embedded in extensive "clutter" of urban environments and may involve well trained individuals exploiting different tactics to execute tasks pertaining to their mission while organizing themselves in such a way that obscures the detection of their cover, access, sources, and responsibilities. The ability to identify adversarial intent of group activities based on the analysis of their physical and interaction behavioral patterns would significantly improve asymmetric counter-insurgency and peace-keeping operations. At present, adversarial intent of group activities are judged by soldiers observing such group activities which often entail significant danger and possibly high false-positive and false-negative rates [1]. Ability to automatically identify adversarial intent via fusion of data from multi-source sensors will yield a unique opportunity to reduce false-alarm rates and facilitate discovery of potentially perilous group activities while helping to shift the balance in peace-keeping operations.

4. Project Motivation

The proposed project was motivated by the realization that numerous group activities may happen under certain circumstances that limits our inclusive visibility of their operations. Since many of such group activities are monitored by surveillance cameras, there is a great need for robust techniques facilitating automatic analysis and interpretation of imagery data associated with such group activities and generating pertinent information about their nature of operations.

For instance, some such group activities might occur inside or within a close proximity to a vehicle/boat/airplane, or at corners of an alley in an urban environment, or at a security bunkers, or behind large obstructions in an urban environment, or at a border location where only partial visibility of group activities is available. An exemplary case is that of the in-vehicle group activity participated by a team of insurgents. The state of behavioral patterns of insurgents in such a situation can be divided into multiple phases: An "arrival phase", where insurgents approach the vehicle from different sides either simultaneously or at different time interval; an "in-vehicle phase", where some interactions among participating insurgents take place inside the vehicle, and a "departure phase", where insurgents depart the vehicle in a certain pattern. By spatiotemporal inferencing and reasoning of such observations, the adversarial intent of such insurgency group activity can be potentially identified. In this project, we are interested to develop robust imagery techniques facilitating better tracking and improve techniques needed for better understanding of In-Vehicle Group Activities (*iVGA*) under partial visibility constraints.

5. Scope of This Feasibility Study

This project has two phases. This report covered our research progress toward the first phase of this project called the feasibility study. The second phase of this project is out scope of this report and therefore is not discussed in the body of this report.

This feasibility study is further divided into two parts, a theoretical feasibility study and an experimental feasibility study. The theoretical study will begin with development of context-based taxonomy and ontology schema for coherent analysis and inferencing of the *iVGA*. One other objective of this theoretical study is to develop suitable adaptive image processing techniques for characterization of *iVGA*.

By taxonomy, we refer to classification of *iVGA* systematically. By ontology, we mean formal representation of knowledge in the form of a collection of concepts within a taxonomy domain and the relationships among such concepts. The ontology-based concepts can be used to reason about involving entities and their relationships within the specified taxonomy domain. In study of the *iVGA*, learning to associate and correlate sub-group activities is the key for being able to differentiate normal behavioral patterns from abnormal ones.

The theoretical effort of this project develops ontological models is intent to cope pattern templates useful for detection, recognition, and discovery of a potential *iVGA* adversarial intent. Another aspect of this preliminary feasibility study is to establish an *iVGA* ontology model based on which an automatic surveillance system can properly relate its observational information discovery (i.e., evidential cues) to recognition of *iVGA* that may pertain an adversarial intent.

6. Research Accomplishments

The research accomplishments of this project are multiple fold. The followings present a summary of our research accomplishments including:

6.1 Development of Space-state Model for Occupants Inside a Vehicle - The space-state allows to recognition normal zones in the physical environment that we expect the normal present of different body part of the passenger, e.g., head, arms, and hands. Figure 1 presents the space-state zones corresponding to normal spaces where a driver head and forearm is expected during normal operation inside a vehicle. Figure 2 illustrates the designated spaces where the hand will be under normal operations inside a vehicle. Figure 3 shows the designated zones where head and hands will be expected during certain operations inside the vehicle. Figure 4 demonstrates the designated spaces for cameras and video camcorders that are determined kinematically sensible (i.e., comfortable) configurations for taking pictures and videoing from inside a vehicle. These space states facilitate inferring and discriminating normal postural configuration of passenger inside vehicle from variance postural configuration that may have security concern inference.

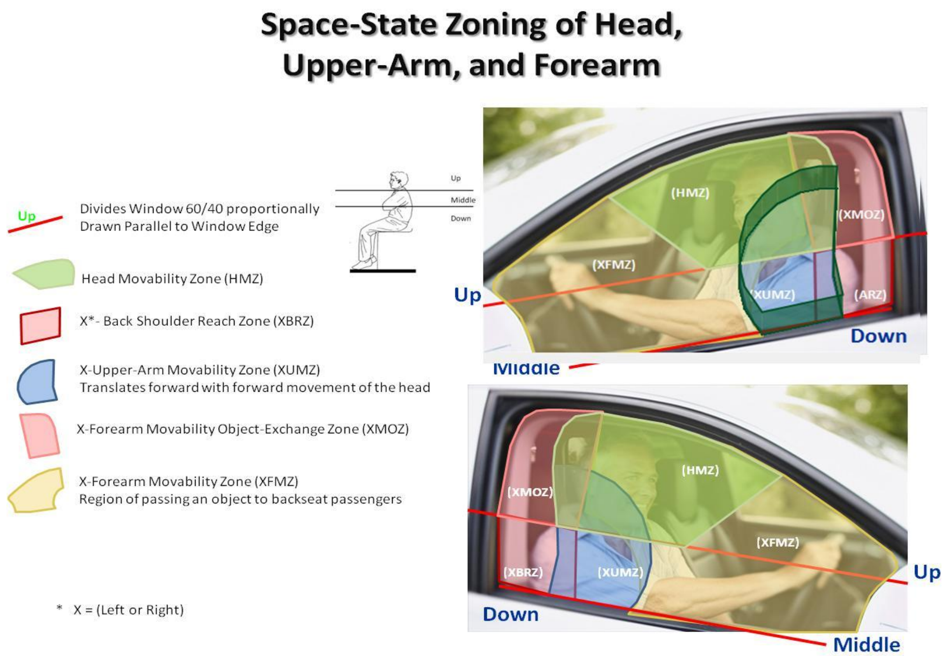

Figure 1. Space-state Zoning of Driver Inside a Vehicle.

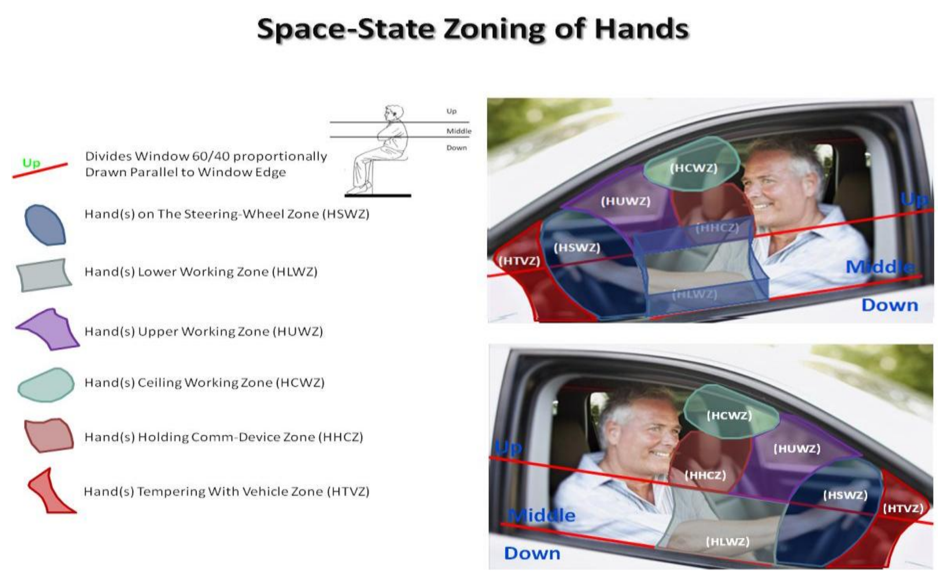

Figure 2. Space State of Driver Hand Inside a Vehicle

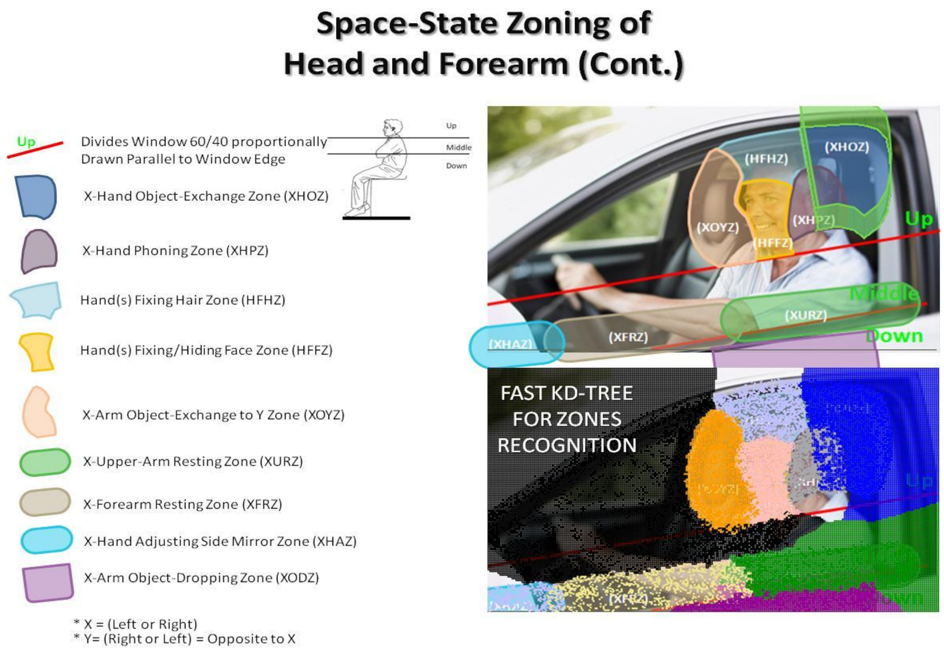

Figure 3. Space-state Zoning of Driver Head and Forearm Inside a Vehicle.

Figure 4. Space-state Zoning of Driver For Special Applications, e.g., taking picture, videoing a scene.

6.2 ***Development of Ontology for Inferring Occupants' Postural Configurations Inside Vehicle*** **-** These newly developed ontology allow to infer certain specific postural configuration of the passenger inside vehicle to certain normal operation inside the vehicle, e.g., adjust rear viewing mirror, turning steering wheel, resting arm against vehicle door window edge, or extending arm outside vehicle through window opening to drop something outside. Figure 5 illustrates kinematic division of an arm with three distinctive parts: Upper Arm, Forearm, and Head. As demonstrate each arm segment at any given kinematic configuration may be described via a tri-state notation. Up, Mid, or Down. This implies that one can use nine degrees of freedom to describe the kinematic postural configuration of an arm. Figure 6 illustrates the kinematic configurations of upper arm relative to the physical world (i.e., vehicle). Figure 7 presents the taxonomy of iVGA. The upper graph in the figure 7 presents the tri-state of vehicle (Parked (i.e., stationary), Arriving, or departing (i.e., leaving). Under each state, the expectation is either the vehicle is occupied or not occupied. If vehicle is found occupied, then, the passengers could be considered as either driver or passengers. The driver is always found inside and in front of vehicle. Whereas, the other passengers may be found sitting inside, and in front of vehicle next to the driver, or sitting in the back seat right behind the frontal driver and passenger. Figure 8 illustrates the taxonomy of whereabouts of the driver and the Passenger(s) if found outside of vehicle. Figure 9 presents an ontology chart of an occupied vehicle with driver and passengers(s) when the driver and passengers head and arms are either visible or invisible. Note that when either head or arm is invisible, no specific deduction is considered. And when the arms are visible, the arms may be either moving or stationary (i.e., not moving). In the latter situation, each state implies its own inferences. Figure 10 summarizes the ontology of a driver head in visible state inside a vehicle. This figure illustrates the inference of driver head while turning, or looking down, looking up, and looking straight ahead.

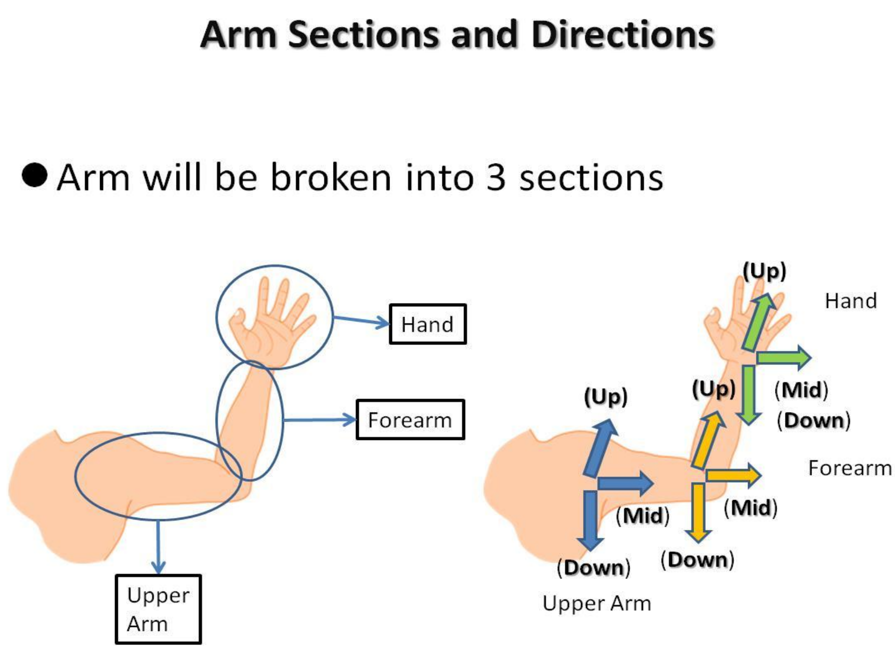

Figure 5. Three-segmented regions of an arm (left illustration) and tri-state of arm kinematic configuration (right illustration).

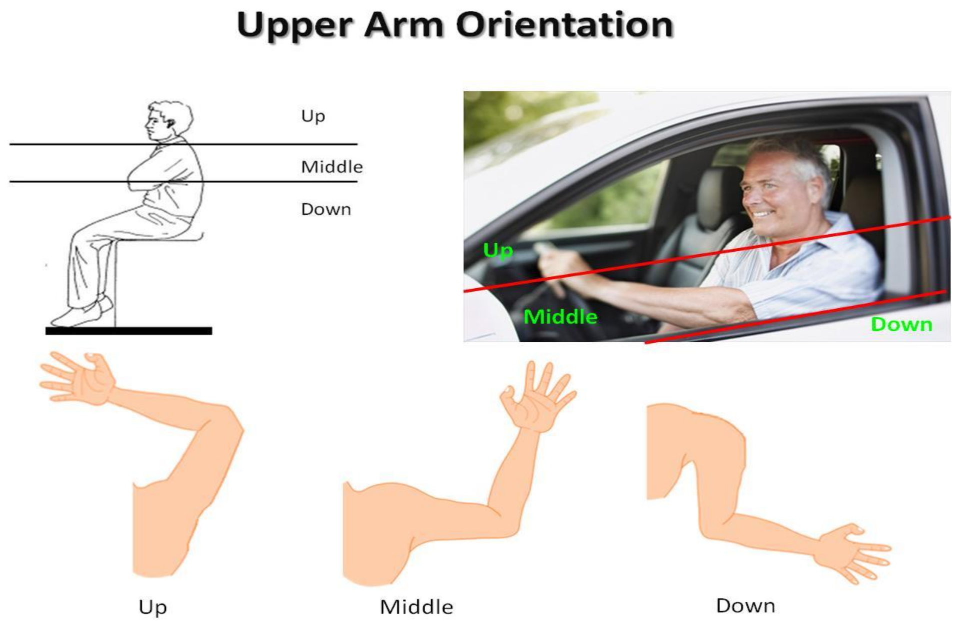

Figure 6. Three Sections of Driver's Side Window, Upper Part, Middle Part, and Down Side. Note the dividing lines are adjustable parameters that can be decided manually for the user or automatically by the image processing algorithms attempt to interpret an *iVGA* scenario.

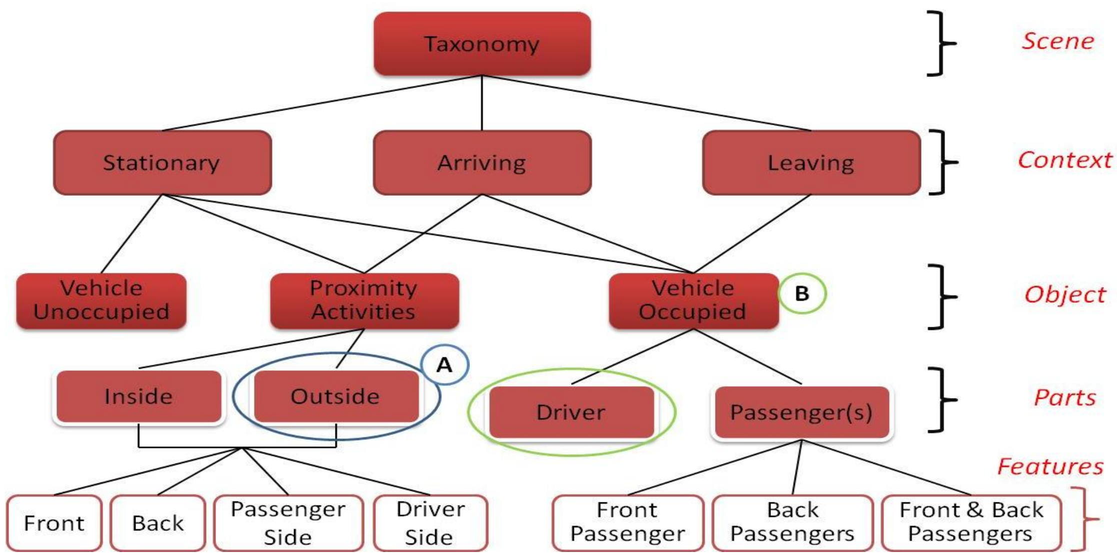

Figure 7. Taxonomy of iVGA – Upper figure shows the tri-state of vehicle (Parked (i.e., stationary), Arriving, or departing (i.e., leaving). Under each state, the expectation is either the vehicle is occupied or not occupied. If vehicle is found occupied, then, the passengers could be considered as either driver or passengers. The driver is always found inside and in front of vehicle. Whereas, the other passengers may be found sitting inside, and in front of vehicle next to the driver, or sitting in the back seat right behind the frontal driver and passenger.

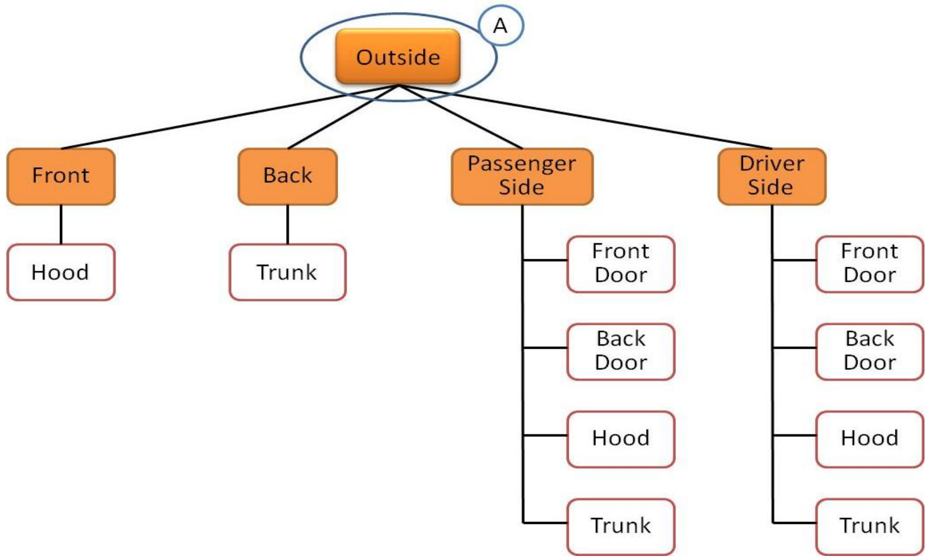

Figure 8. Taxonomy of whereabouts of the driver and the Passenger(s) if found outside of vehicle.

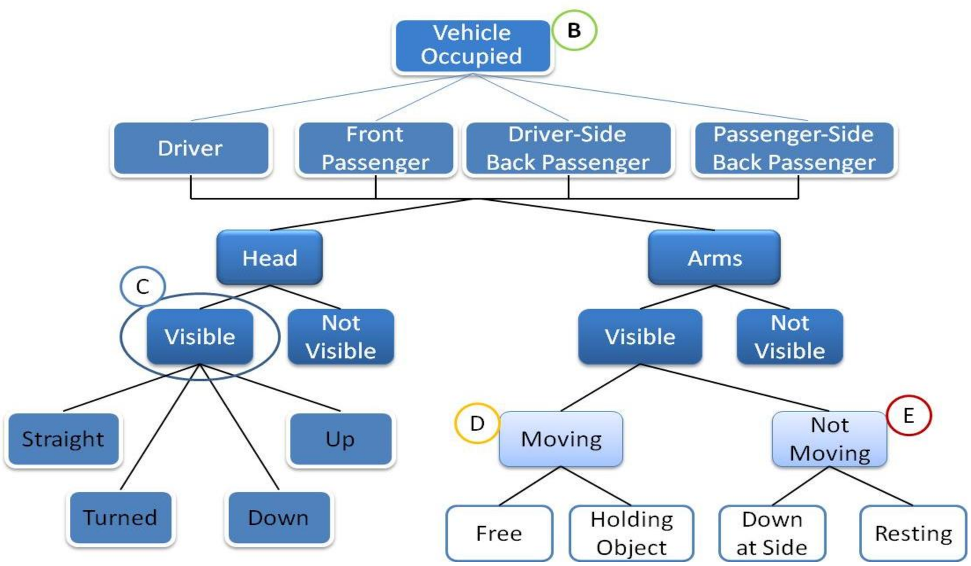

Figure 9. Ontology of an occupied vehicle with driver and passengers(s) where theirs head and arms are either visible or invisible. Note that when either head or arm is invisible, no specific deduction is considered. And when the arms are visible, the arms may be either moving or stationary (i.e., not moving). In the latter situation, each state implies its own inferences.

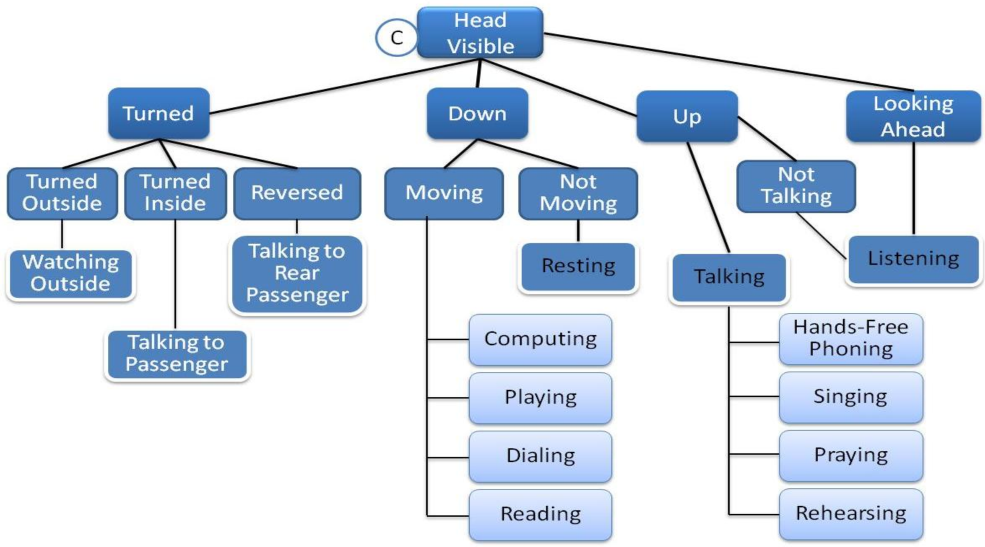

Figure 10. Ontology of a driver head in visible state and implication of driver head while turning, or looking down, looking up, and looking straight ahead. When the head is turned, it may be turning inside or outside or "reversed". The latter implies to a situation where the driver may be talking to a passenger located in the back seat. The deductions from the head configuration in state of down and up are plausible inferences that may require further interrogation.

Figure 11 presents the ontology of the driver arm is moving in two states: (1) free state and (2) holding state. In the free-state, the moving arm may be pointing upward to downward. Implication of either state is modeled in the left leaf of chart in the figure 11. In the state of holding an object, the action of object holding may have accomplished either by one hand or two hands. Each state, again, has its own implication as illustratively demonstrated by the right leaf of chart in the figure 11. For instance, an action to phoning is involved with movement of an arm while holding an object (e.g., a phone) and positioning it somewhere around the head. This composition of the arm moving can be concluded that a phone call is being placed once such actions are observed. On the other hand, when an object is held by two hands, the hands may be found either below the chest area, around the head area, or above the head area. Each situation has its own implication as captured by the right leaf of chart in the figure 11. For instance, a situation where two hands of the drivers are found moving toward his face area, may indicate that perhaps, the driver is either adjusting his/her eyeglasses, adjusting his/her makeup, or possibly attempt to hiding his/her face from getting recognized. In normal situation, for instance, it is not customarily for a person to use two hands around his/her face area. This does not mean it is not absolutely impossible, but rather expresses its out of normality. Moreover, when the hand are found above the face and over the head, one possible logical explanation of that is that it's likely that individual is fixing his/her hair or adjusting his/her hair cover. Note that ontology describe here can be decisive in terms of explanation normality of postural states of an individual inside a vehicle. Many exceptions can occur that are not deemed significant to be include in the ontology charts. In this project, we are after modeling normal situations, and assumption is that anything outside this normality regime may be anomalous and subject to further examination and scrutiny through this analysis. Figure 12 illustrates the ontology of situation a drive arm are found visible but not moving. Figure 13 presents the notations developed for describing the postural states of arms.

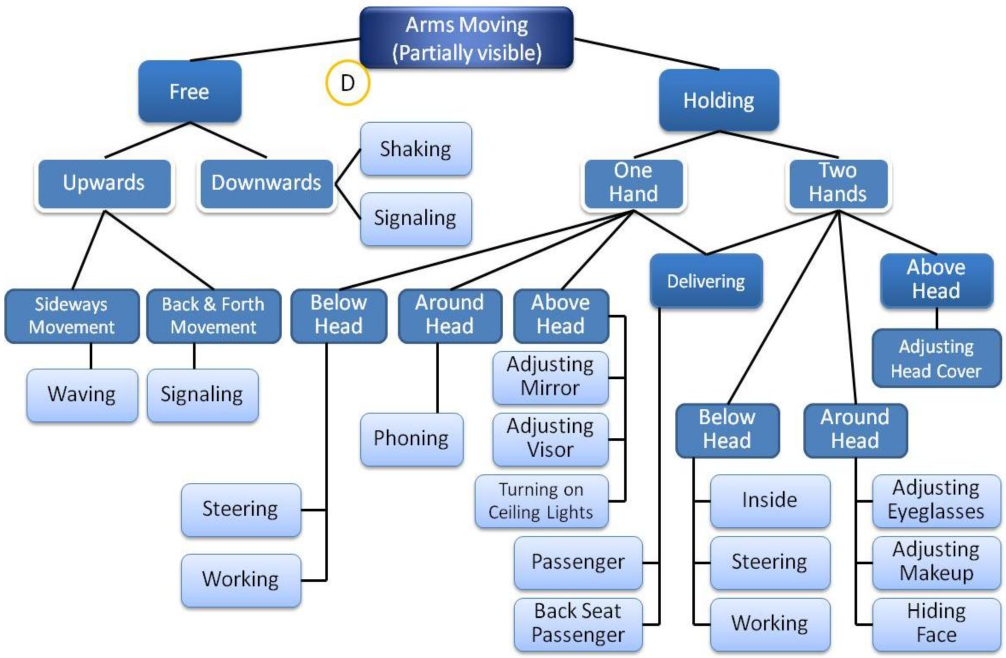

Figure 11. Ontology for state of the driver arm moving.

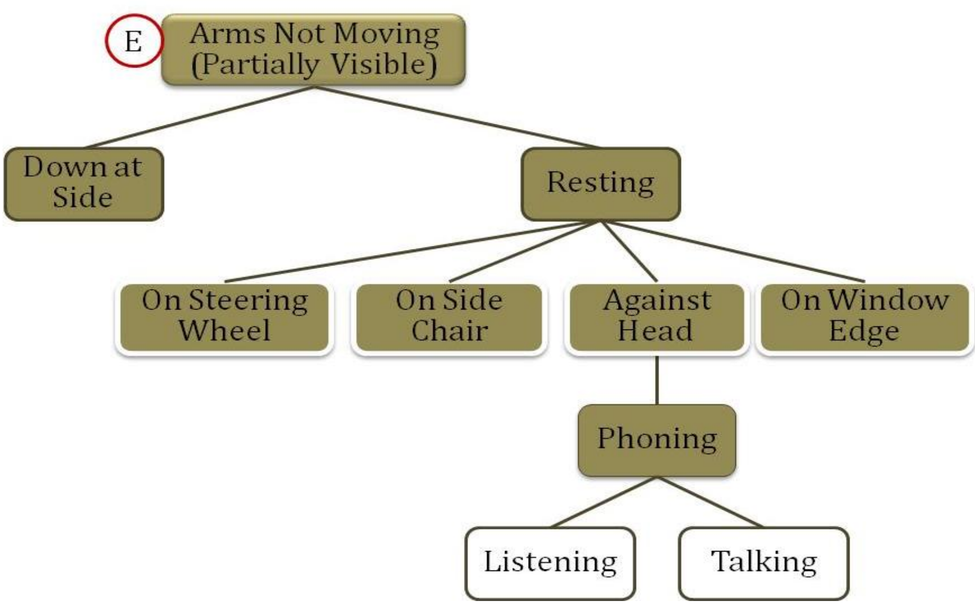

Figure 12. Ontology of a driver arm detected but found not moving.

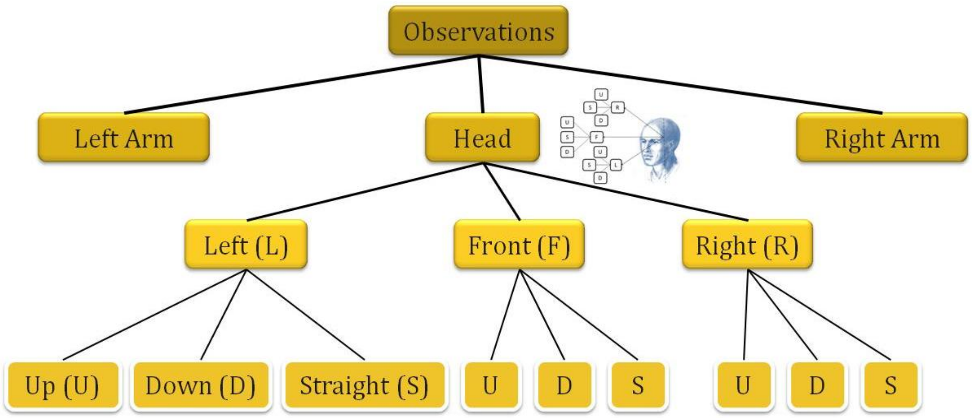

Figure 13. Ontology notations for describing the postural states of arms.

Figure 14 presents an Arm Gesture Temporal Modeling. A typical arm movement can be described by a three temporal states including: Resting (R), Transitioning (T), and Stroking (S). Each state is realized with a time period. Arms prefer to come to rest after each movement. This is due to conservation of energy and improving extra load bearing of the biological states. Arms in extended configuration typically require more energy for their control and hence a tiring kinematic configuration for the human being. However, arms are rest, consumed less energy and cause no additional burden for the brain to monitored their control. A much elaborated arms movement may involve a combination of transitioning, namely, moving from one state or another in order to accomplishment some action, or then it typically comes to rest via a rapid stroke to ease pressure to muscles controlling the arm in the extended kinematic configurations. Figure 15 presents the notations for describing arms and hands postural/gestural configuration.

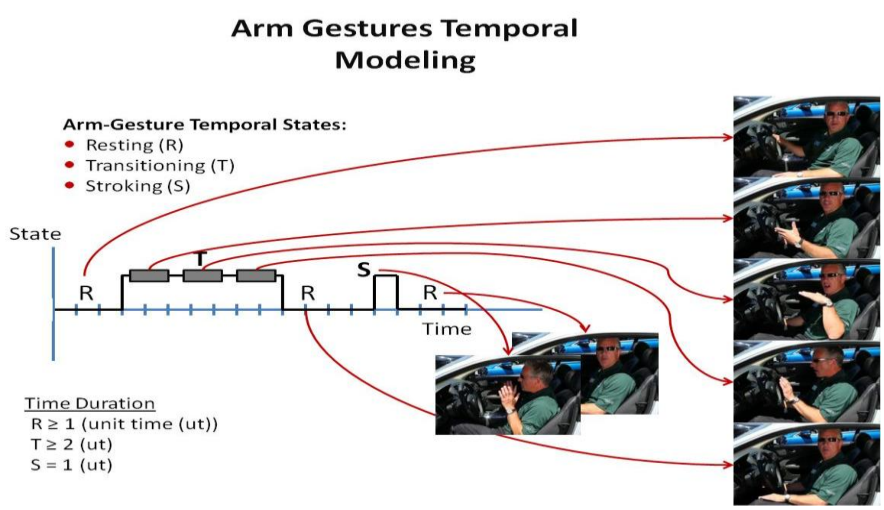

Figure 14. Arm Gesture Temporal Modeling Via Three States: Resting (R), Transitioning (T), and Stroking (S).

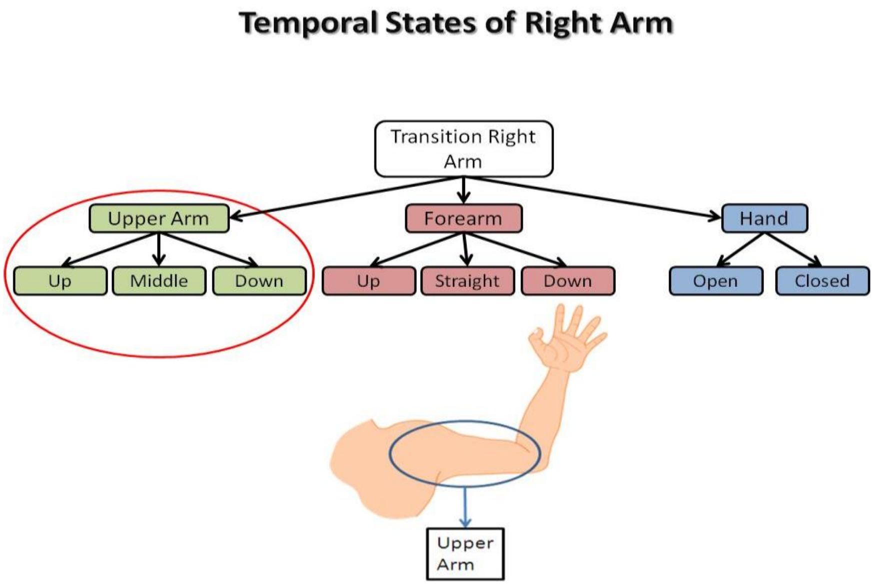

Figure 15. Ontology for notations for arms and hands postural/gestural configuration.

Figure 16 presents the ontology of right upper arms of the driver in three states of up, middle, and down configuration. A typical person uses his/her arm to do specific functions while seating in the driver seat. For example, adjust rear viewing mirror is highly likely is done using right arm than left arm. When an upper arm is observed, that observation alone, if confirmed and found in either up, middle or down kinematic configuration, then, the possible functions that person may be performing is captured in the state of model of the ontology chart in the Figure 16. This ontology model excludes operations that are not kinematically feasible or unlikely to be permissible to be performed with the upper arm while it is in any other unspecified states. Figure 17 presents ontology developed to note the forearm postural/gestural configurations. Figure 18 presents the ontology of right upper arms of the driver in three states of up, middle, and down configuration. Similar to the description provided for the Figure 16, figure 18 also captures the essence of forearm kinematic configuration and its implication when detected in either of those known states. For example, this ontology reveals that for adjusting rear viewing mirror, for example, the forearm must be seen in an upper configuration. This also supports the previous ontology for the right upper arm configurational requirement for the same function. Namely, for adjusting rear viewing mirror, both the right upper arm and the right forearm must be detected in the upward configuration, otherwise any violation of this requirement does not grant the perception of this operation. Moreover, when the right upper arm and the forearm are found in the upward configuration and toward the middle upper center of vehicle toward the front this grant the perception of reaching for the rear viewing mirror and attempting to modify it. This is a knowledge-based postulation that may require further reinforcement via supplementary examination and scrutiny of this operation.

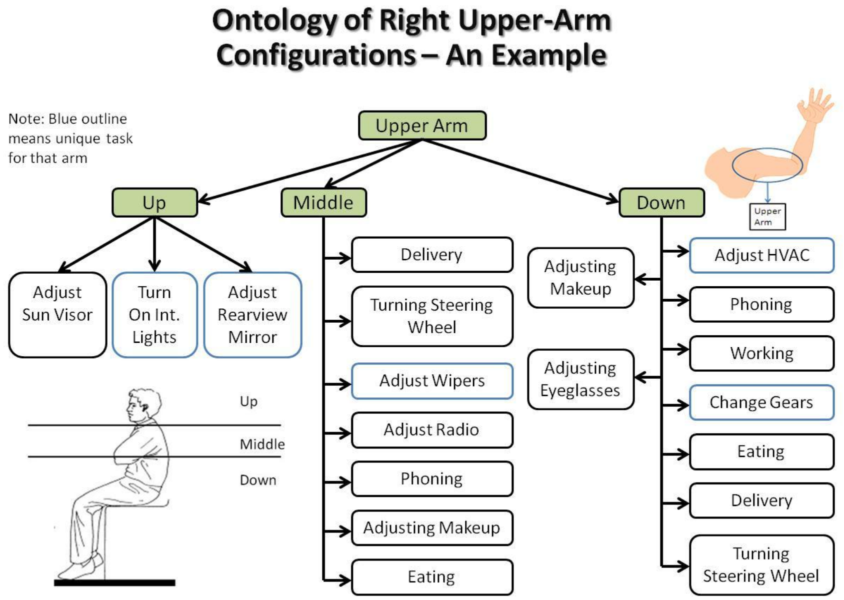

Figure 16. Ontology of Right Upper-Arm in three states of Up, Middle, and Down.

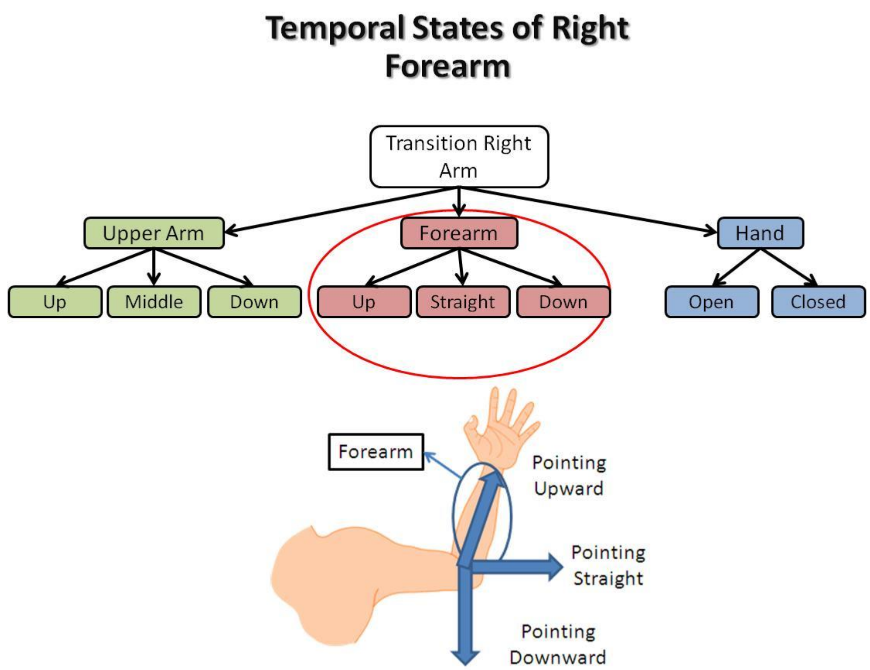

Figure 17. Ontology for notations for Forearm postural/gestural configuration.

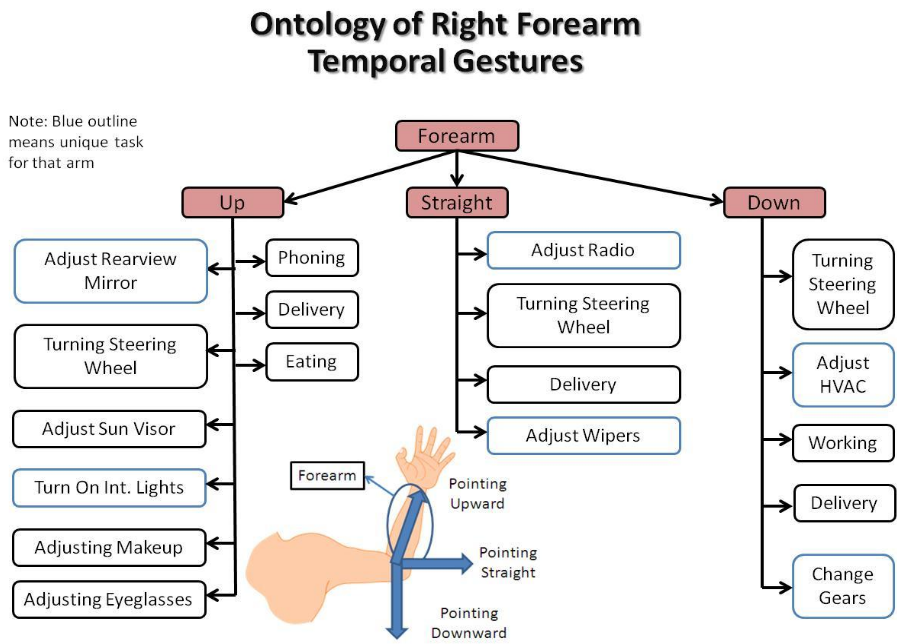

Figure 18. Ontology of Right Forearm in three states of Up, Middle, and Down.

Figure 19 presents the ontology of hand. Figure 20 presents the ontology of temporal configuration assumable by hand in its open or closed gestural configuration. Human hands are very dexterous and used for performing tremendous number of functions particularly for manipulation objects. Inside the confined space of a vehicle, right hand and left hand are unlikely to be duplicating each other normal function as far as the basics of operations of vehicles are concerned. For instances, turning on/off the interior lights of the vehicle are highly likely performed by the right hand because of its spatial closeness to the ceiling lights of the vehicle that are always centrally situated inside the vehicle, instead of using left hand. Such operation is not a matter of preference, but rather again, a matter of energy conservatism that is in inherent attributes of biological systems. On the other hand, many actions of an object manipulation may be handled by either hand. Under this ontology model the manipulative configurations of the hand is collectively considered as "working" configuration with no further refinement because of kinematic complexity of hands that makes their configurational identification ambiguous.

Similarly for the left arm, we developed the complimentary ontology for the left upper arm, left forearm, and left hand. These ontologies are demonstrated in Figures 19 thru 26 and are self-explanatory.

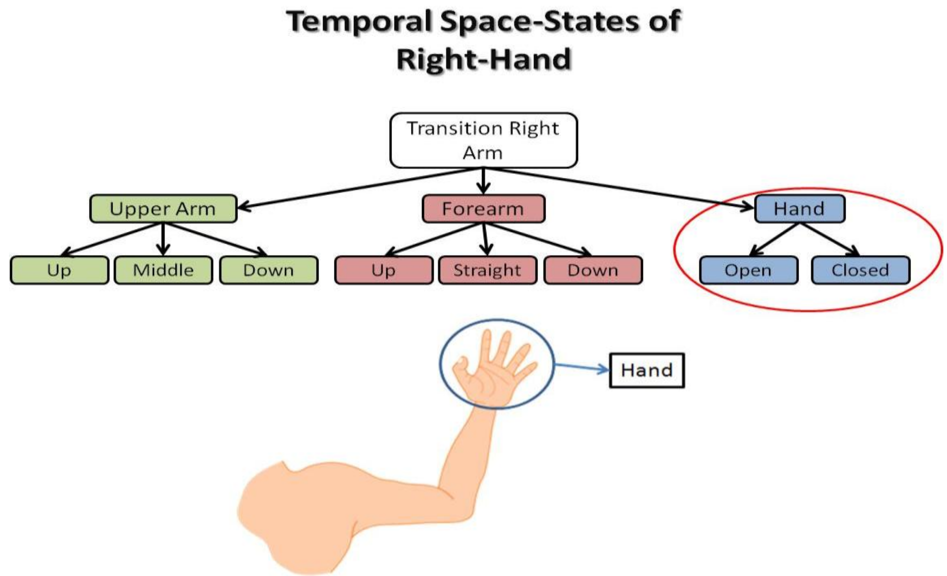

Figure 19. Ontology for notations for hand gestural configuration.

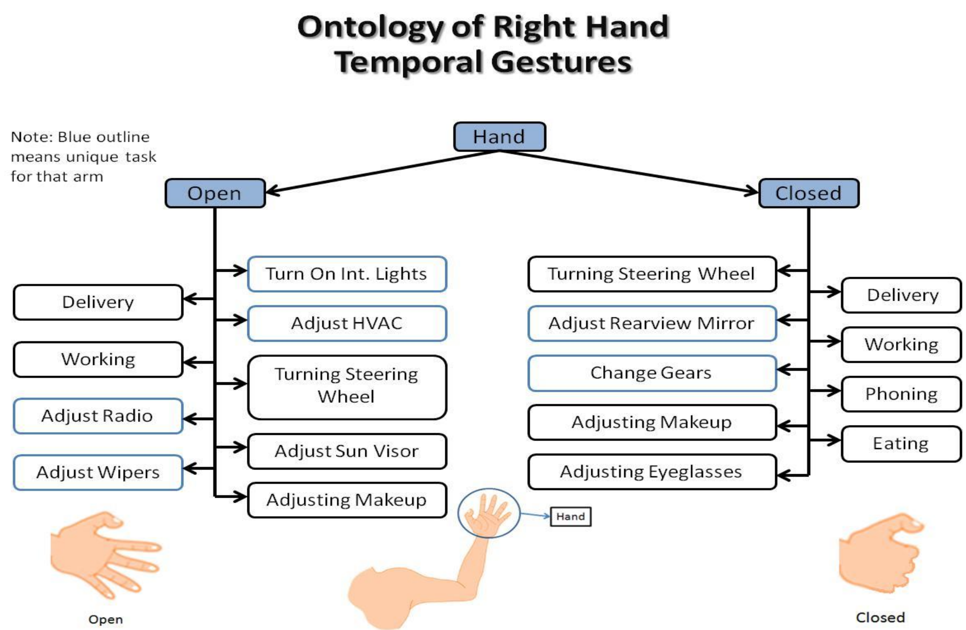

Figure 20. Ontology of Right Hand in two states of open and close.

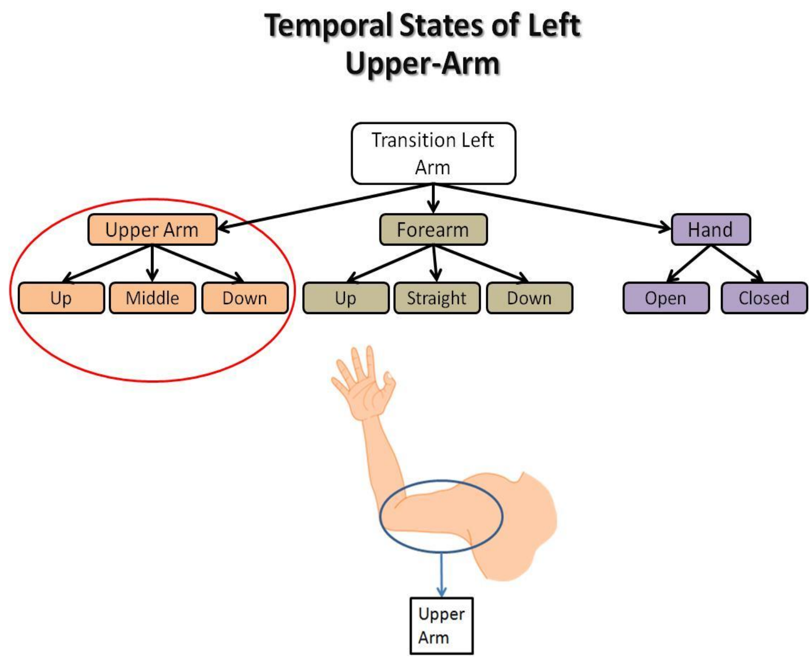

Figure 21. Ontology of left upper arm.

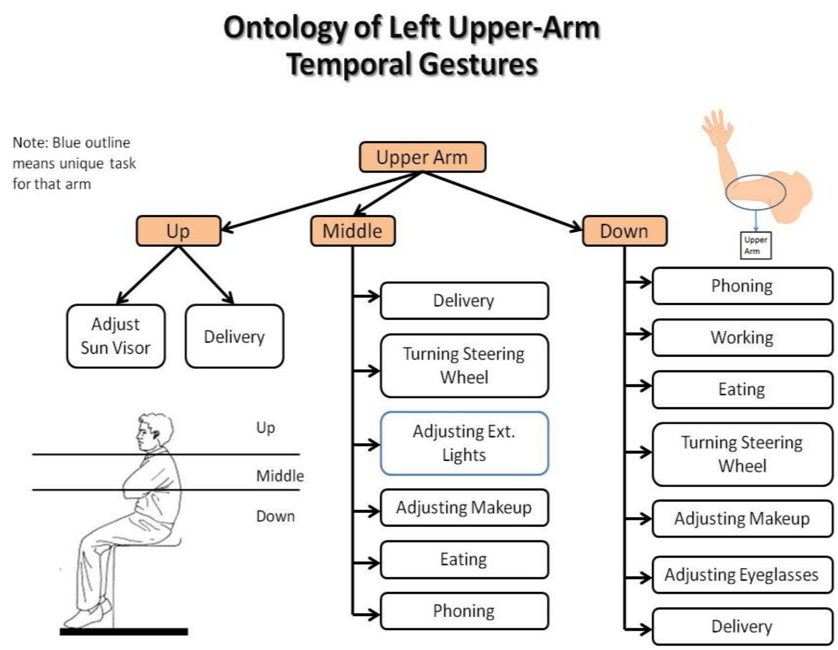

Figure 22. Ontology of left upper arm for three states of Up, Middle, and Down.

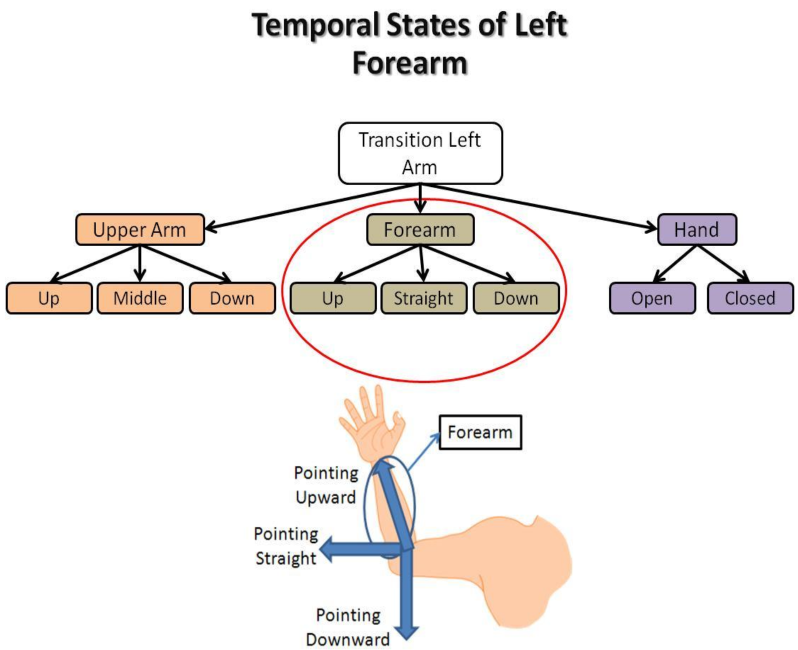

Figure 23. Ontology of Left Forearm.

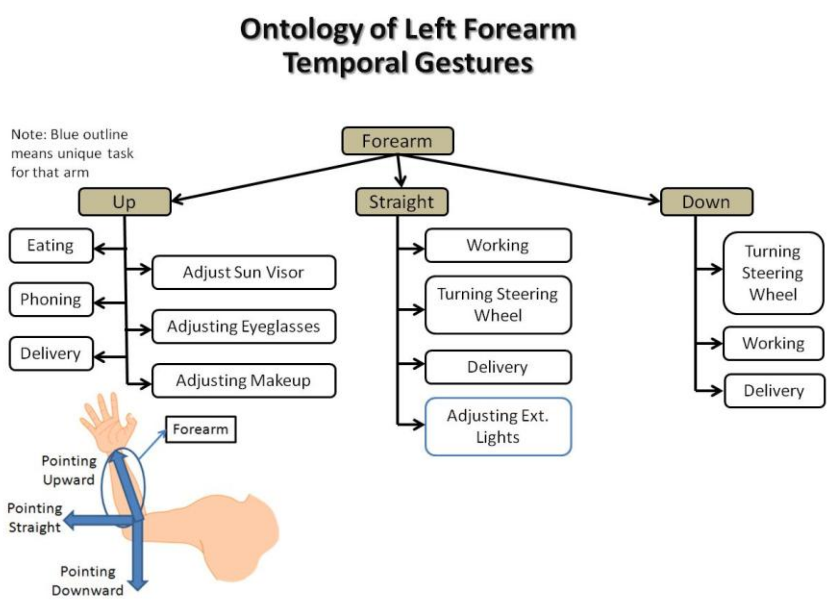

Figure 24. Ontology of Left Forearm for three states of Up, Middle, and Down.

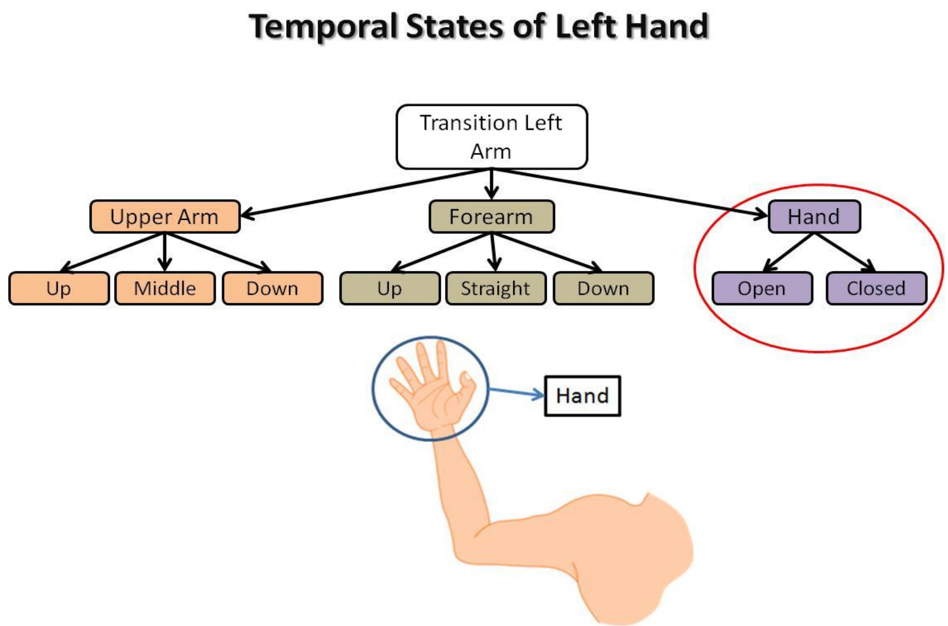

Figure 25. Ontology of Left Hand.

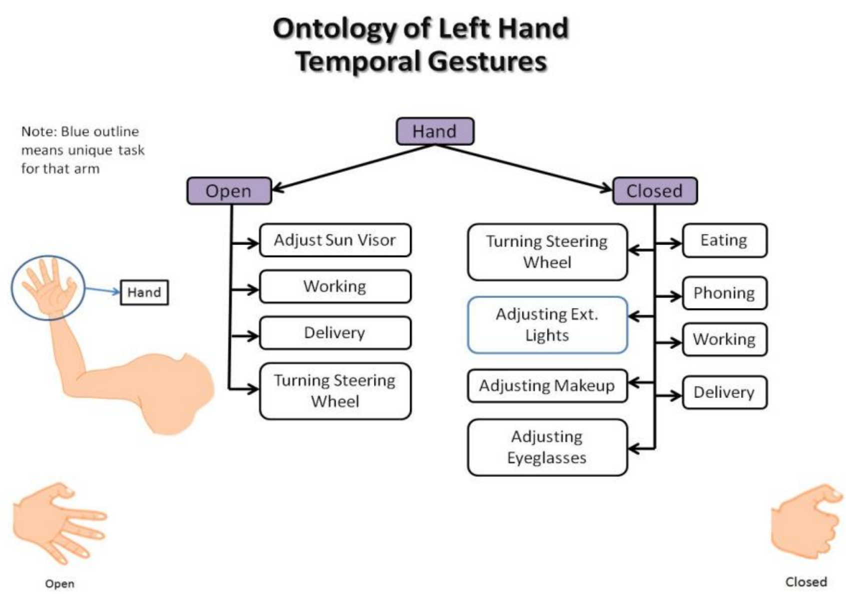

Figure 26. Ontology of left hand for two state of open and close.

6.3 Development of Arms Kinematic Configuration Admissibility (AKCA) Decision Trees

The newly developed ontology for left and right arm only consider arm postural configuration that are kinematically achievable. However there are many combination of arm kinematic configurations that not achievable. Categorically, we call this category of kinematically unachievable arm postural configurations as "kinematically admissible arm postural configurations". These inadmissible arm postural configurations are traceable and we managed to capture these singularities an AKCA decision tree as shown in Figure 27.

To arrive at the AKCA decision tree, we employ a technique called ID3. In decision tree learning, ID3 (Iterative Dichotomiser 3) is an algorithm invented by Ross Quinlan [1] used to generate a decision tree from a dataset. ID3 is the precursor to the C4.5 algorithm, and is typically used in the machine learning and natural language processing domains.

The ID3 algorithm begins with the original set Sas the root node. On each iteration of the algorithm, it iterates through every unused attribute of the set $S$and calculates the entropy $H(S)$(or information gain $IG(A)$) of that attribute. It then selects the attribute which has the smallest entropy (or largest information gain) value. The set S is then split by the selected attribute (e.g. age < 50, 50 <= age < 100, age >= 100) to produce subsets of the data. The algorithm continues to re-curse on each subset, considering only attributes never selected before. Recursion on a subset may stop in one of these cases:

- Every element in the subset belongs to the same class (+ or -), then the node is turned into a leaf and labeled with the class of the examples
- There are no more attributes to be selected, but the examples still do not belong to the same class (some are + and some are -), then the node is turned into a leaf and labeled with the most common class of the examples in the subset
- There are no examples in the subset, this happens when no example in the parent set was found to be matching a specific value of the selected attribute, for example if there was no example with age >= 100. Then a leaf is created, and labelled with the most common class of the examples in the parent set.

Details of the ID3 is summerized in Figure 28. Throughout the algorithm, the decision tree is constructed with each non-terminal node representing the selected attribute on which the data was split, and terminal nodes representing the class label of the final subset of this branch. The AKCA decision trees can serve as look up table and facilitate verification and validation of a kinematic configuration of arms, namely, helps in acceptance or rejection of a kinematic arm configuration suggested by the image processing algorithms detecting passenger's arm kinematic configuration. Note that, certain arm configurations are inadmissible - kinematically- speaking, e.g., hand is dexterous, but its degree of freedom becomes limited under certain kinematic configuration of its arm. Figure 29 illustrates another decision tree generated using ID3 algorithms that summarized the logical for determination of which hand gestural configuration is admissible.

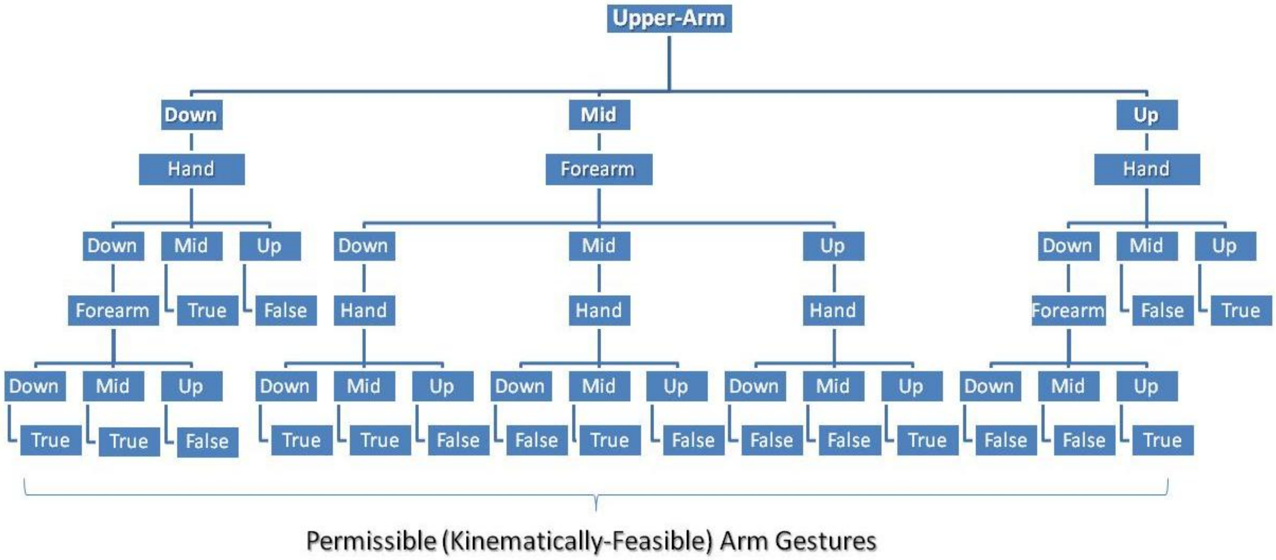

Figure 27. A generalized decision-tree of kinematically admissible arm postural configurations.

```
ID3(D, Target, Atts)
returns: a decision tree that correctly classifies the given examples

variables
D: Training set of examples
Target: Attribute whose value is to be predicted by the tree
Atts: List of other attributes that may be tested by the learned decision tree

create a Root node for the tree
if D are all positive  then Root ← +
else if D are all negative then Root ← –
else if Attrs = ∅ then           Root ← most common value of target in D
else
      A← the best decision attribute from Atts
      root ← A
      for each possible value v_i of A
            add a new tree branch with A=v_i
            Dvi ← subset of D that have value v_i for A
            if Dvi = ∅  add then leaf ← most common value of Target in D
            else add the subtree ID3( Dvi, Target, Atts-{A} )
```

Figure 28. Pseudo-code description of ID3 decision-tree algorithm.

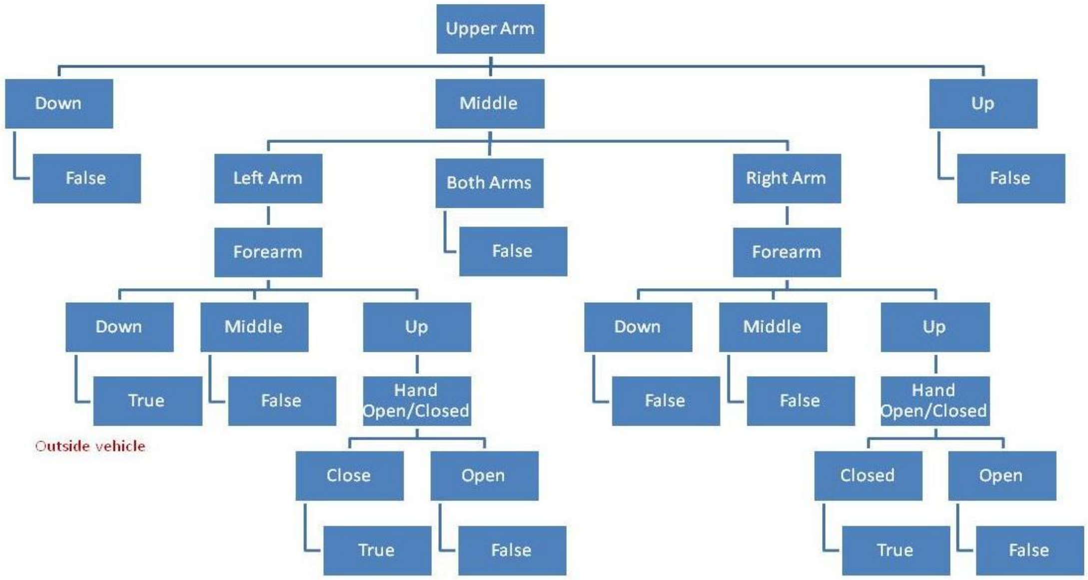

Figure 29. A generalized decision-tree of kinematically admissible Hand-down gestrual configurations.

6.4 Development of New Algorithms for Vehicle Windows Detection - In this feasibility study, we considered situations when the vehicle under surveillance is immobile and activities are taking place inside the vehicle. Given a perspective view of the vehicle, the objective is to determine that section of vehicle corresponding to windows with visibility to see through the vehicle. There are many variety of vehicle are available. In this project, we consider a few vehicles as illustrated in the following figures. Windows are an integral part of vehicle and all upper portion of most commercially available vehicle are surrounded by windows with different shape, and transparency. The camera observing the vehicle is assumed to be located above the height of the driver inside the vehicle and therefore, the camera is looking forward with slope through the window to observe the activities of vehicle occupants. Primarily, we considered the driver since each vehicle is driven with at least one driver. Other vehicle occupants are optional and arbitrary. Therefore, and foremost, we are interested to characterize the activity of the driver through a window orientation revealing his/her actions. Figure 30 presents the results of several vehicle edge detection Image Processing (IP) algorithms developed for detection my lines representing the structural configuration of the vehicle. A typical commercial vehicle has many surface features and contains many reflective regions because of curvature of its body surfaces, parts, and corners. A robust edge-detector algorithm must isolate the most strongest edges representing the geometrical shape of the vehicle without jeopardizing the structural integrity of the vehicle and resulting of uncharacteristic edge formations. Figures 31 and 32 demonstrate different edge detector algorithms developed and tested for the purpose. Figures 33 thru 35 demonstrate the next step of this process for detection of front window and side windows of the vehicle.

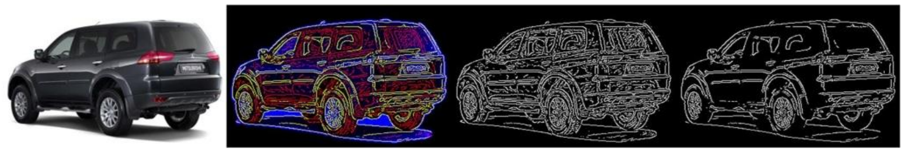

Figure 30. Result of first image processing technique for vehicle edge detection.

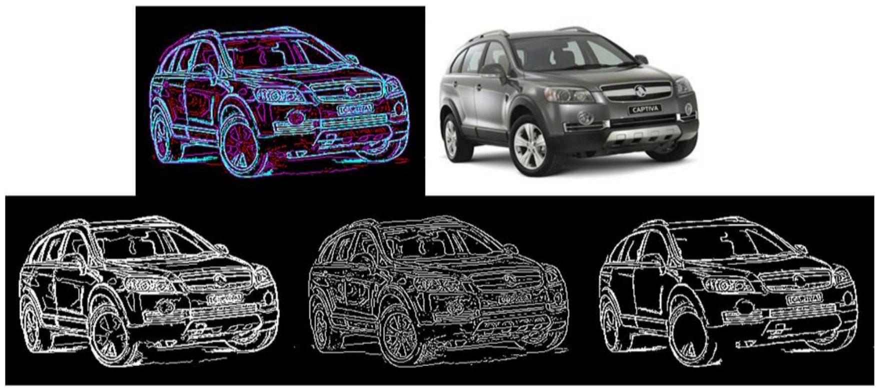

Figure 31. Results of another image processing technique for vehicle edge detection.

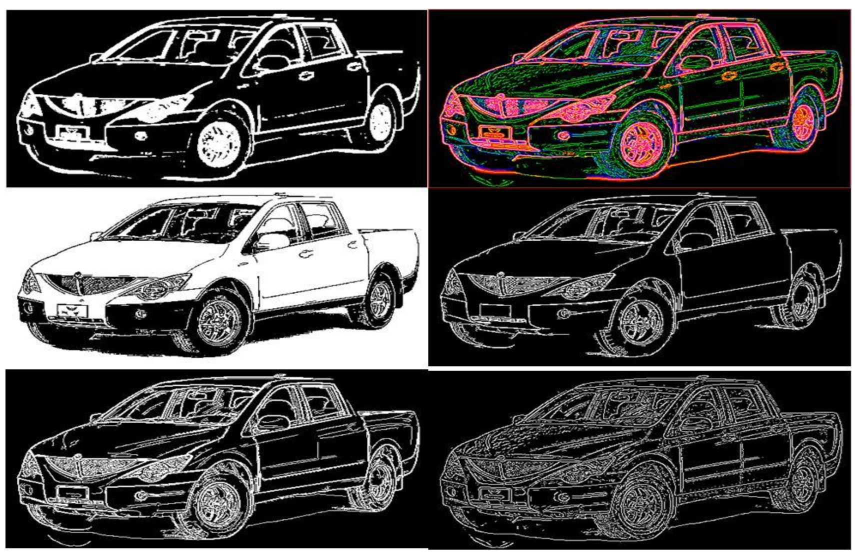

Figure 32. Results of third image processing technique for vehicle edge detection.

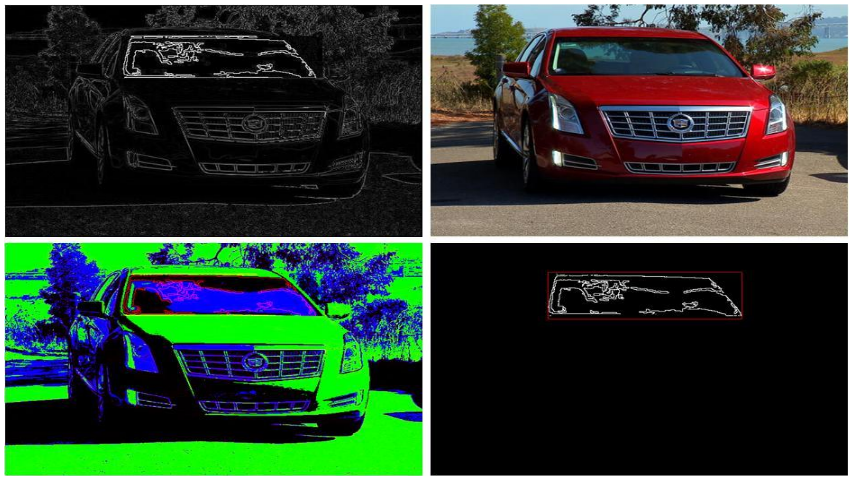

Figure 33. Results of fourth image processing technique for detection of vehicle frontal window.

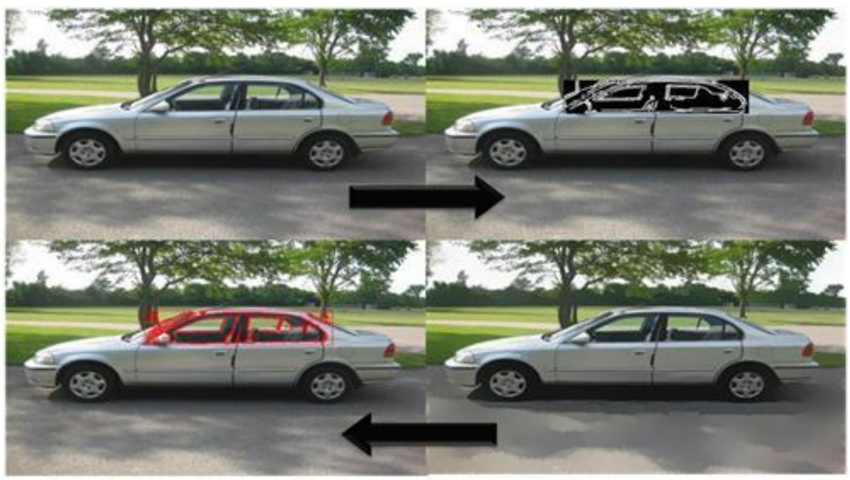

Figure 34. Results of fifth image processing technique for detection of vehicle side window.

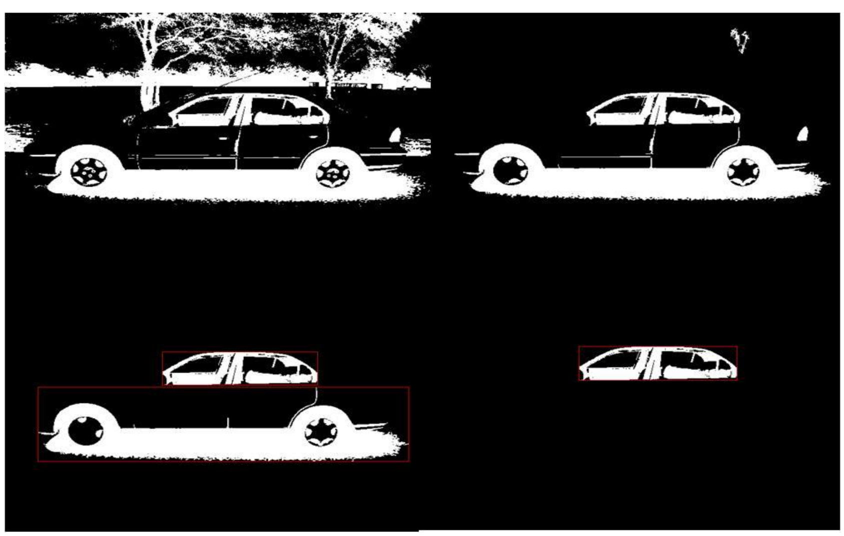

Figure 35. Results of sixth image processing technique for detection of vehicle side window from original vehicle binary image.

6.5 Development of New Algorithms for Human Skin Detection - Upon detection of vehicle window area, the next step is to isolate head, left, and right arms. The common denominator of these three body parts is the skin. Therefore, we embarked to develop a robust skin detector based on YCrCb color space. Figure 36 illustrates the regime of Cb vs. Y'-Cb chart, and Cr vs. Y'-Cr chart selected for segmentation of skin colors. These two chart was developed heuristically and by examination of hundreds of different color skins from variety of ethnicity groups with different skin colors. Any point in the YCrCb that can be projected into the red regions of charts in the Figure 37 corresponds to skin point that our algorithm recognizes it. Figure 37, a picture from Wikipedia website used for testing of this newly developed skin detector. Prior to applying this filter, noise variation of the origin image needs to be suppressed. We tried different digital filter for this purpose including Median, Bilateral, and Circular filters that we found are the most effective for suppressing RGB color noises and resulting more consistent and smoother skin colors. Figure 38 illustrates the effectiveness of our newly developed skin detector for variety of people with different ethnicity background and skin colors with partial shading effects. Selection of the noise suppression filter is critical, since overdoing of this option suppressed equally many facial features that appear relative small in the pictures. Figure 39 demonstrated the strength of this skin color detector for people from the same ethnicity but with different size, shading, and partial obstructions and occlusions. Figure 40 shows the effectiveness of this approach for a face that is partial, but completely darkened by the shadow.

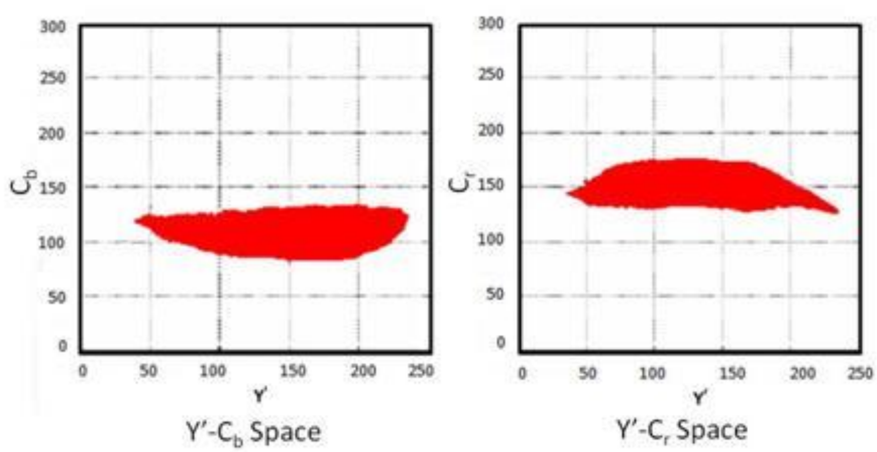

Figure 36. Charts of Y-Cr and Y-Cb spaces for skin color detection.

Figure 37. Results of skin detector: Original (Upper Left), Large Median Filter+Skin Detector (Upper Right), Bilateral Filter+Skin Detector (Lower Left), Circular Filter+Skil Detector (Lower Right) – Courtesy of Wikipedia.

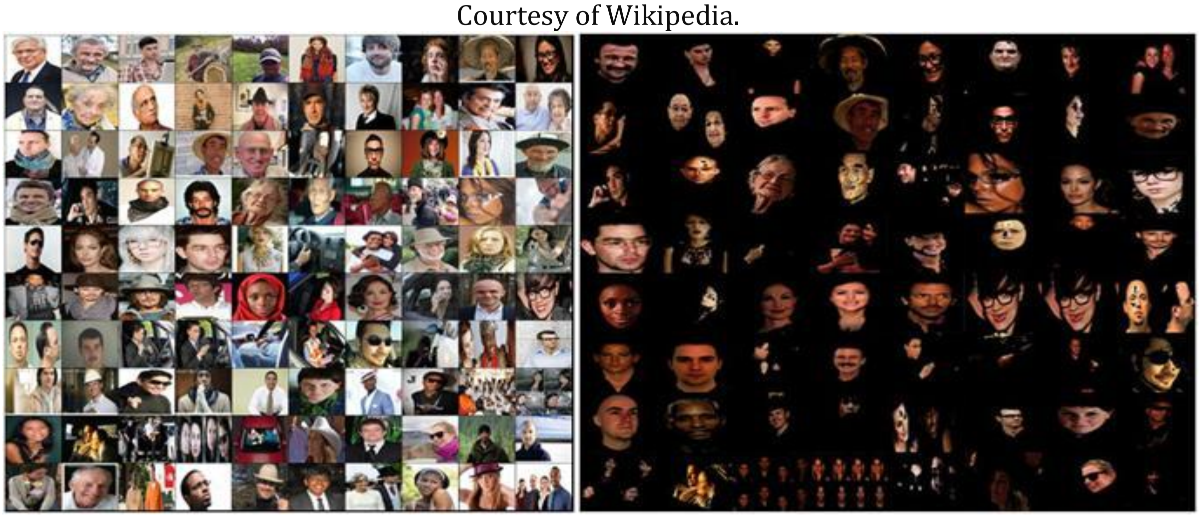

Figure 38. Results of different Samples from variety of different ethnicity groups.

Figure 39. Another Result of skin detection techniques for variety of large to small faces.

Figure 40. Results of skin detection techniques for a bi-modal image – partly lighted up, partly shadowed. Note the hat is detected a part of the skin.

6.6 Development of A New Algorithm for Isolation of Human Body Part From Dynamic Videos - The skin detector was also tested on some YouTube downloaded videos demonstrating unstructured activities inside vehicles. Figures 41 and 42 illustrated the robust of this newly developed skin detector for real-time video processing applications. Note that the computational efficiency of this skin detector is high and it can be readily applied for real-time video processing. In these two demonstrated video, we were about to perform 25 frames per second this option for image size of 320x240 pixels. The videos demonstrated in the figures 41 and 42 contain significant overshadowing effects and activities are rather dynamic. The video in the figure 42 only shows arms/hands movement for assembling of a holder object that is mounted inside of the vehicle for support of an iPad.

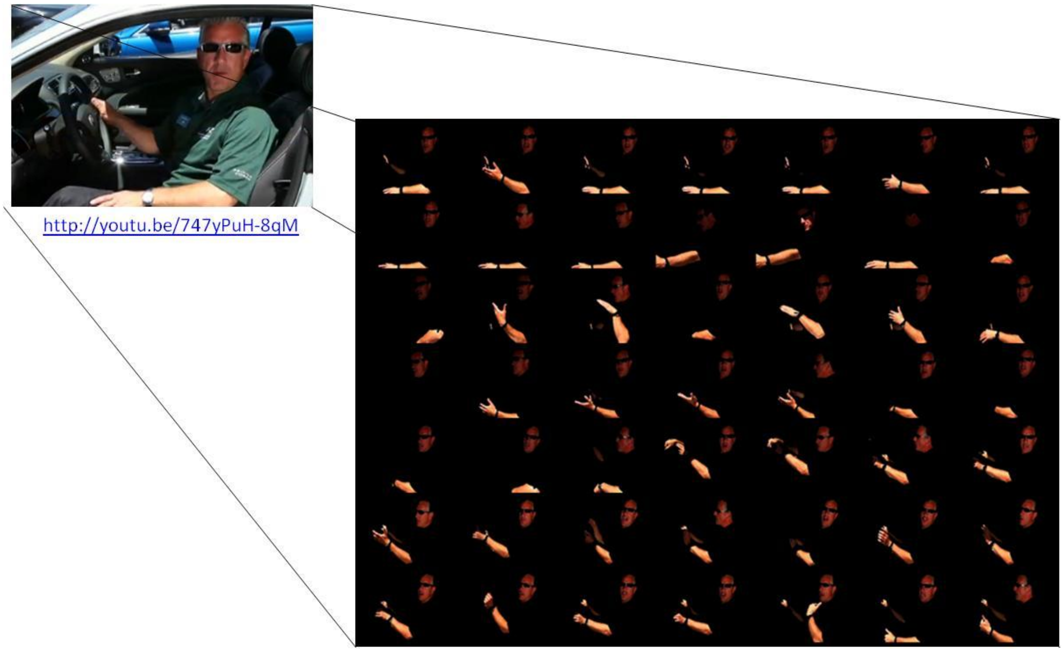

Figure 41. Results of Skin detectors based on a YouTube video demonstrating an in Vehicle Activity.

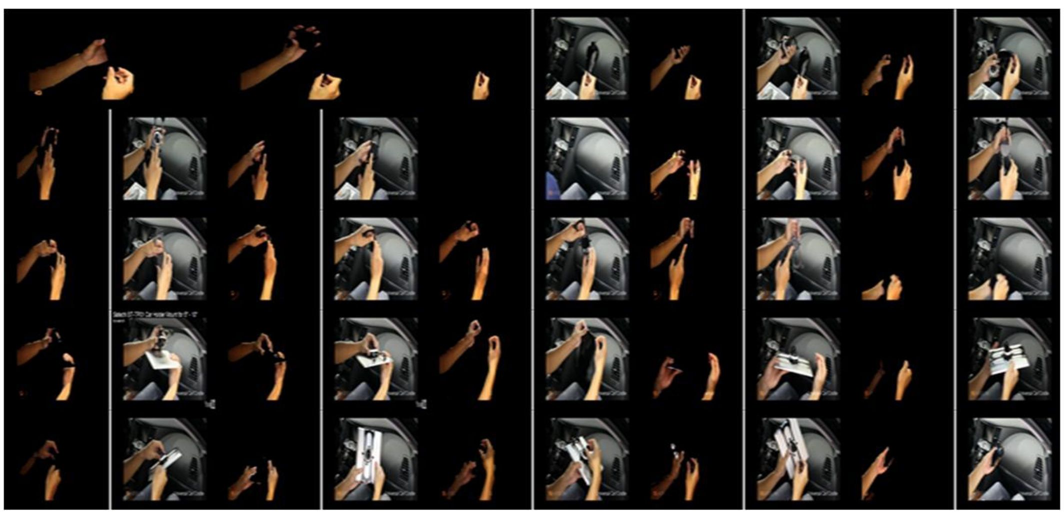

Figure 42. Results of Another Example of Skin detectors based on a YouTube Video Demonstrating An in Vehicle Activity.

***6.7 Development of A New Algorithm for Tracking Human Body Parts Tracking* -** Upon detection of skinned body parts of an occupant, the next step is to isolate each body part and identify their kinematic configuration, namely, their orientation. We use a straightforward logic to isolate these parts. The head correspond to that blob that satisfy certain aspect ratio and most likely located above other body parts. The left and right arms are differentiated by associating the nearby blobs two form most likely representative of the left and right arm. If camera is observing through the driver side, the left arm is considered that blob is the below the head, meets certain aspect ratio, and it is the larger of the other major blob below the head. Using this logically, automatically, the second major blob located below the head is considered as the right arm. Figures 44-49 show the results of tracking of the body parts in different YouTube *iVGA* videos. Figures 50 and 51 presents other YouTube iVGA video processed for the purpose of testing and training of *iVGA* image processing algorithms.

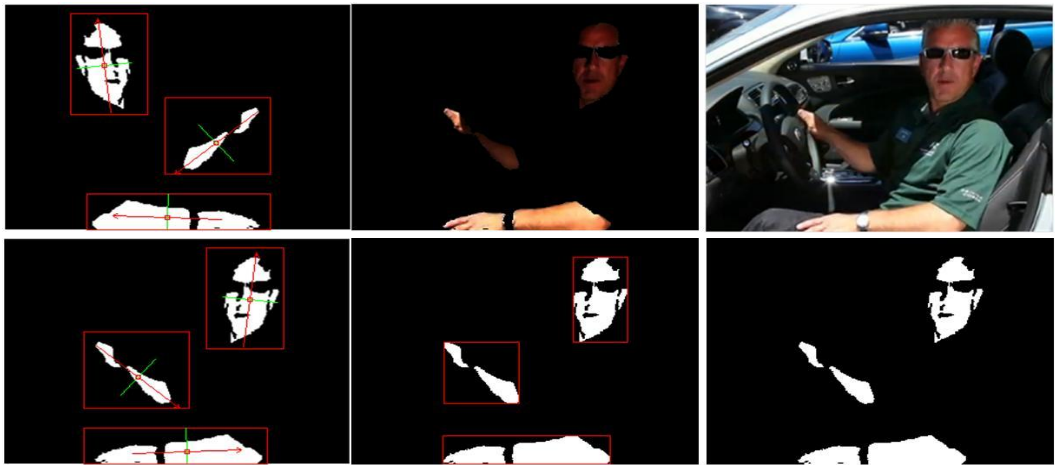

Figure 43. Results for detection, isolation, and characterization of orientation of head and arms.

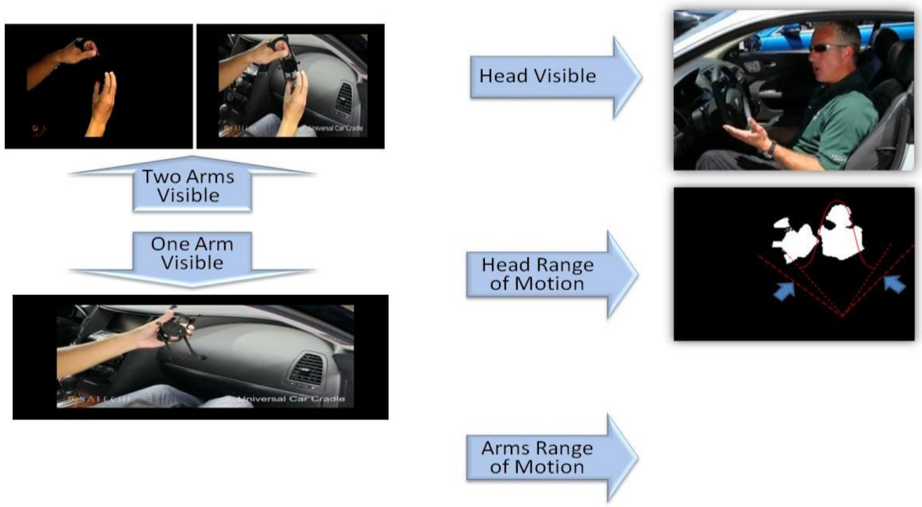

Figure 44. Results for tracking of body parts through sequential body parts.

Tracking Head, Arms, and Hands

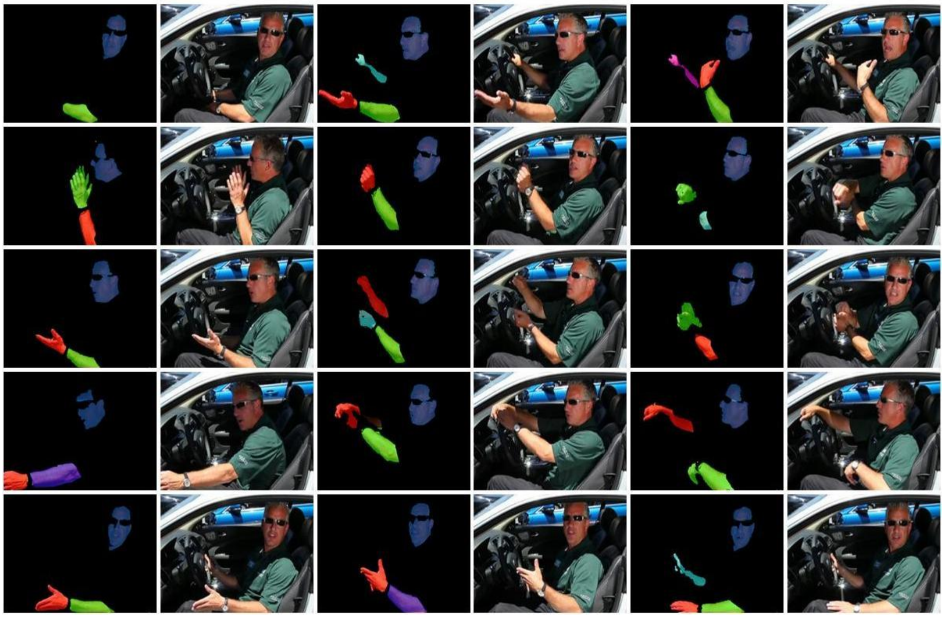

Figure 45. Results for tracking of body parts through sequential body parts for YouTube Video 1. This video is the courtesy of YouTube.

Tracking Working Hands

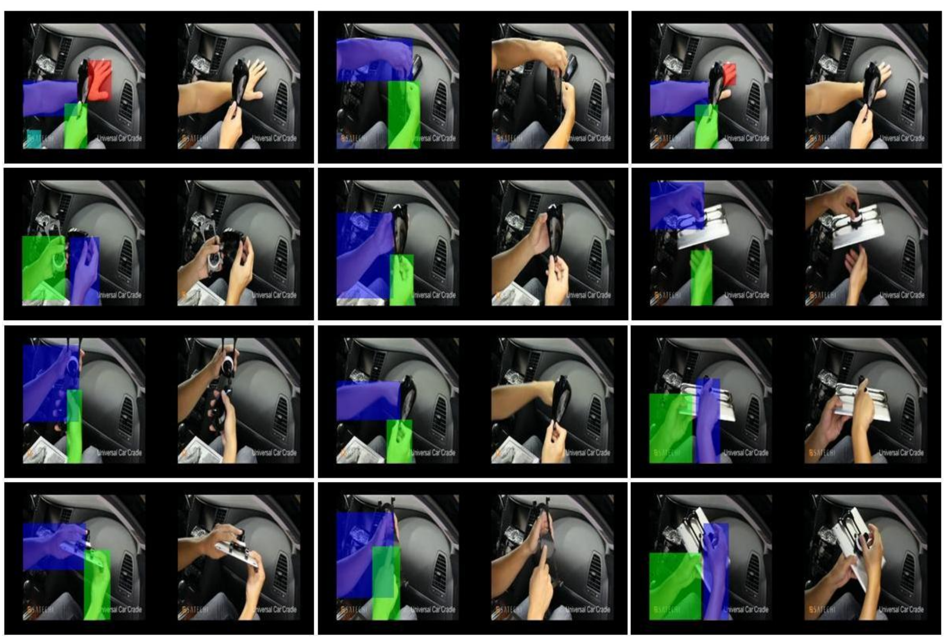

Figure 46. Results for tracking of body parts through sequential body parts for YouTube Video 2. This video is the courtesy of YouTube.

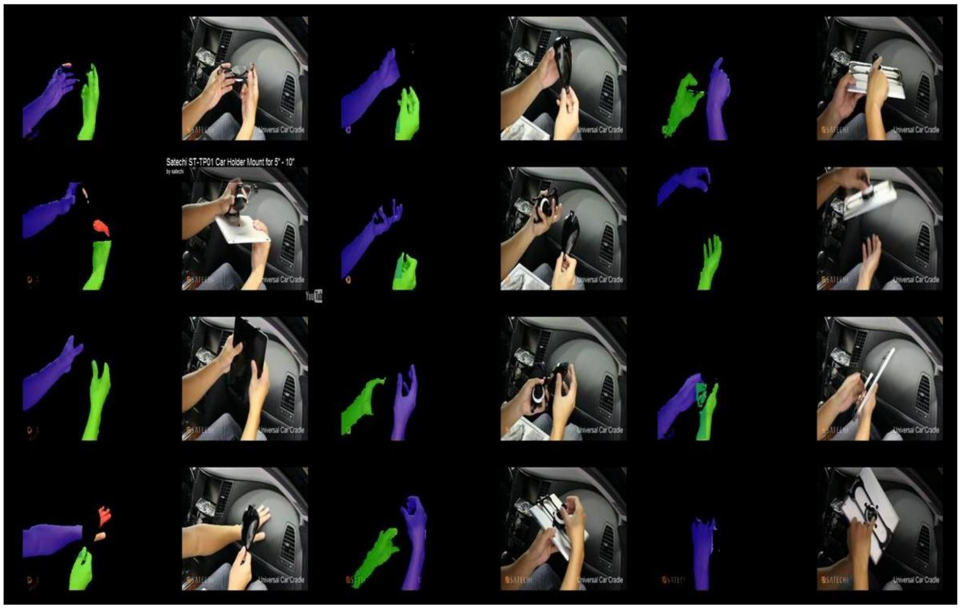

Figure 47. Results for tracking of body parts through sequential body parts for YouTube Video 2. This video is the courtesy of YouTube.

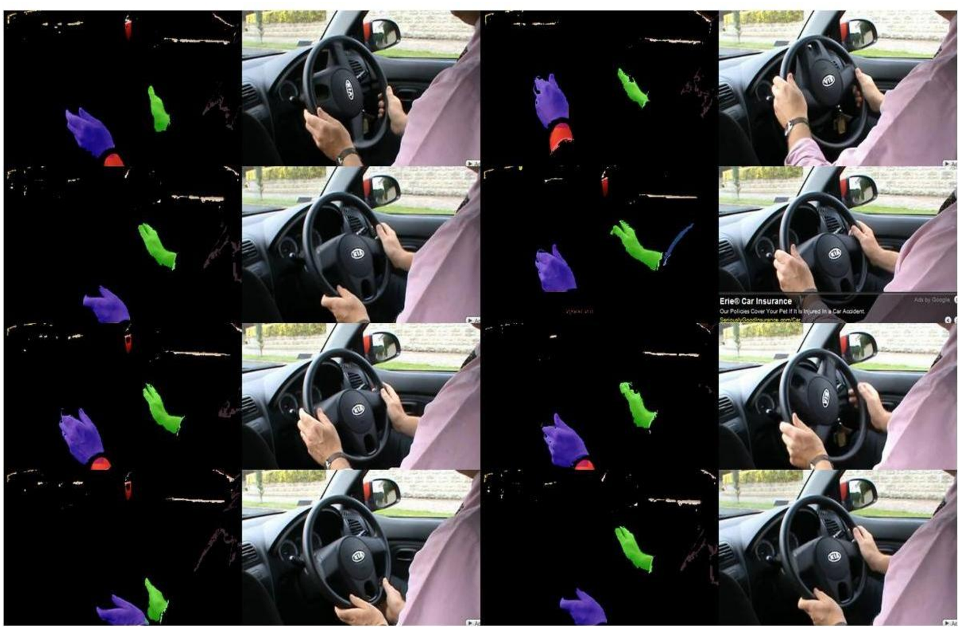

Figure 48. Results for tracking of body parts through sequential body parts. This video is the courtesy of YouTube.

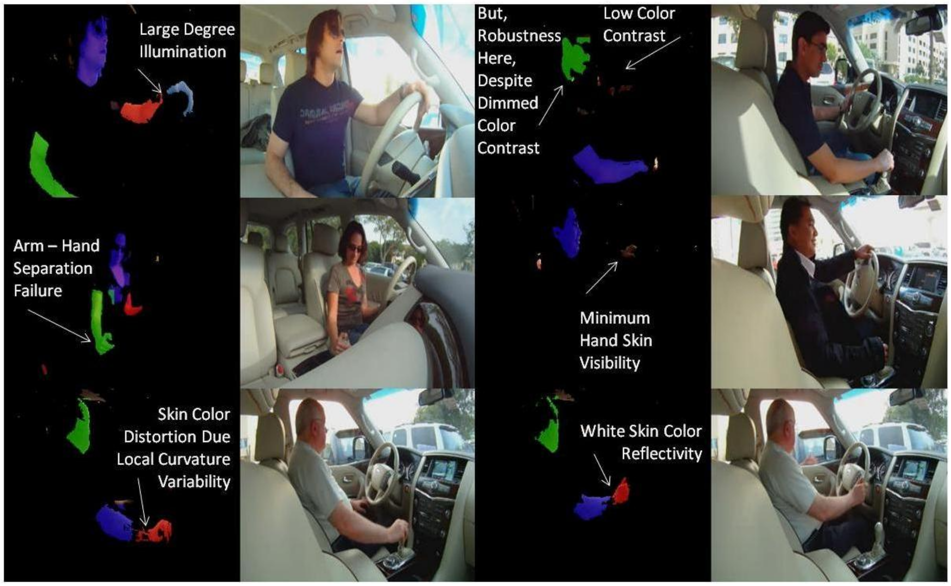

Figure 49. Results new image processing algorithm on different YouTube Videos. Each pair of images demonstratea the situation where tracking of body parts is failed. These videos are the courtesy of YouTube.

http://youtu.be/u0CpJEW5eGE
(0:30-0:40)

http://youtu.be/jDIiUHMcj3w
(0:30-2:00)

http://youtu.be/hhmiJzI5caA

http://youtu.be/Lxa6lF-l13c
(arms/steering wheel)

http://youtu.be/WB6mp3SV_Ww

http://youtu.be/aLqsrb_8G5w
(short video, sleeping)

http://youtu.be/xHJY8rs5JaY
(0:26-0:33, sleeping)

http://youtu.be/Ex3nBWI5F_w

http://youtu.be/qlOQKjkeS6U

Figure 50. A sample of different YouTube videos used for testing and evaluating the newly development *iVGA* algorithms.

http://www.youtube.com/watch?v=WZir-dx5GYo&feature=share&list=PLF024002DBB7D5D94   http://youtu.be/HWEbSX8XT0I   http://youtu.be/1XPejoh6tc0

http://youtu.be/mBX4q0X-bvM

http://youtu.be/VR91PNKlXVo (people smoking, talking)

http://youtu.be/_-z8515pJFQ (convertible, man putting on hats)

http://youtu.be/eyL-SjuJ8p4

http://youtu.be/hU4rxRpyPZQ (side view, door open, long video)

http://youtu.be/4NfMdUs29uw (3:05-4:45, side view, talking)

http://youtu.be/CV5wl2oATn0

http://youtu.be/76C8hATI_Sg (1:25-2:40, side view, seated)

http://youtu.be/747yPuH-8qM   http://youtu.be/0SynvJKXgYI

http://youtu.be/hhCQsVLwREk   http://youtu.be/747yPuH-8qM   http://youtu.be/Lxa6lF-I13c

http://youtu.be/5Nb8iirGe6c   http://youtu.be/XoxksSIJOxc   http://youtu.be/Gb13dJP9bMo

Figure 51. Another sample of different YouTube videos used for testing and evaluating the newly development *iVGA* algorithms.

6.8 Development of A New Algorithm for Initial Detection of Human Body Parts Based on SURF Image Processing Detection - To detect motion we developed a fast and robust algorithm based on a technique called SURF. SURF stands for Speeded Up Robust Features is a robust local feature detector, first presented by Herbert Bay et al. in 2006 [3]. It can be used in computer vision tasks like object recognition or 3D reconstruction. It is partly inspired by the SIFT descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images. SURF uses an integer approximation to the determinant of Hessian blob detector, which can be computed extremely quickly with an integral image (3 integer operations). For features, it uses the sum of the Haar wavelet response around the point of interest. Again, these can be computed with the aid of the integral image. A summed area table is a data structure and algorithm for quickly and efficiently generating the sum of values in a rectangular subset of a grid. In the image processing domain, it is also known as an integral image. It was first introduced to computer graphics in 1984 by Frank Crow for use with mipmaps. In computer vision it was first prominently used within the Viola–Jones object detection framework in 2001. However, historically, this principle is very well known in the study of multi-dimensional probability distribution functions, namely in computing 2D (or ND) probabilities (area under the probability distribution) from the respective cumulative distribution functions [4]. Moreover, the summed area table can be computed efficiently in a single pass over the image, using the fact that the value in the summed area table at (x, y) is just:

$$I(x,y) = i(x,y) + I(x-1,y) + I(x,y-1) - I(x-1,y-1)$$

Figure 52 demonstrates adaptive SURF features learning in real-time. The SURF algorithm is implemented in IRIS image processing library (IRIS is a complete and fast image processing library developed by PI for robotics image processing applications). The IRIS maintains a host of 512 SURF features from a typical video stream. From the first few images of video, SURF reported features are sorted and memorized. Any subsequent features are then compared in real-time against these 512 SURF features. Those features that have the same SURF properties are disregarded since they can be assumed to have belonging to the background. However, the new features that the algorithm can find with no feature match are considered as newer features that may contain the foreground object. IRIS uses an exclusive algorithm to discriminate the new features and eliminated those weak features that unlikely representing foreground. The newer feature found having significant content belonging to foreground are stored in a different learning stack and used for fast testing against future video imagery frames. Figure 53 illustrates the results of this latter image processing techniques for separation of foreground objects from memorized background objects.

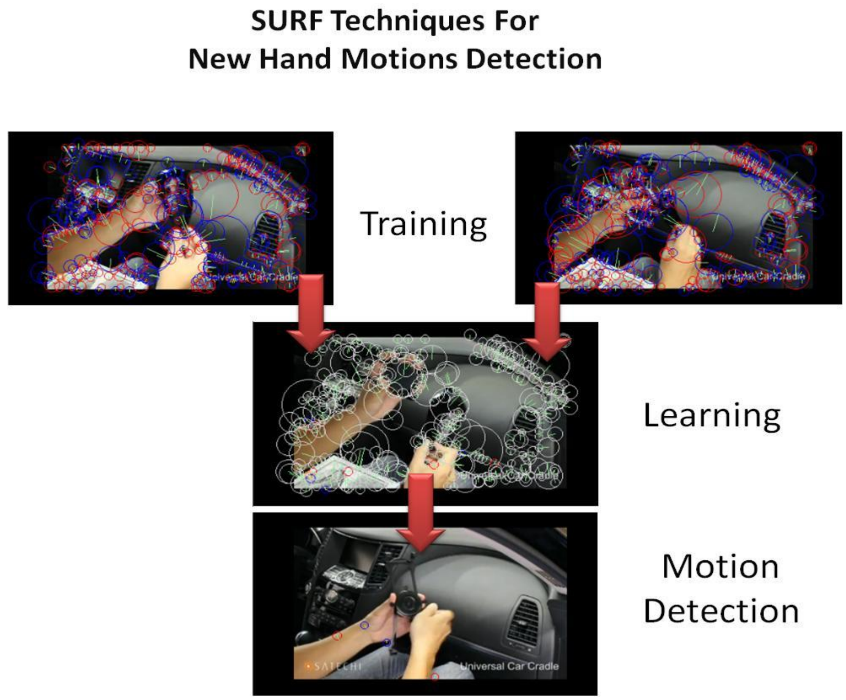

Figure 52. Results for SURF algorithm in IRIS application for detection of moving body parts.

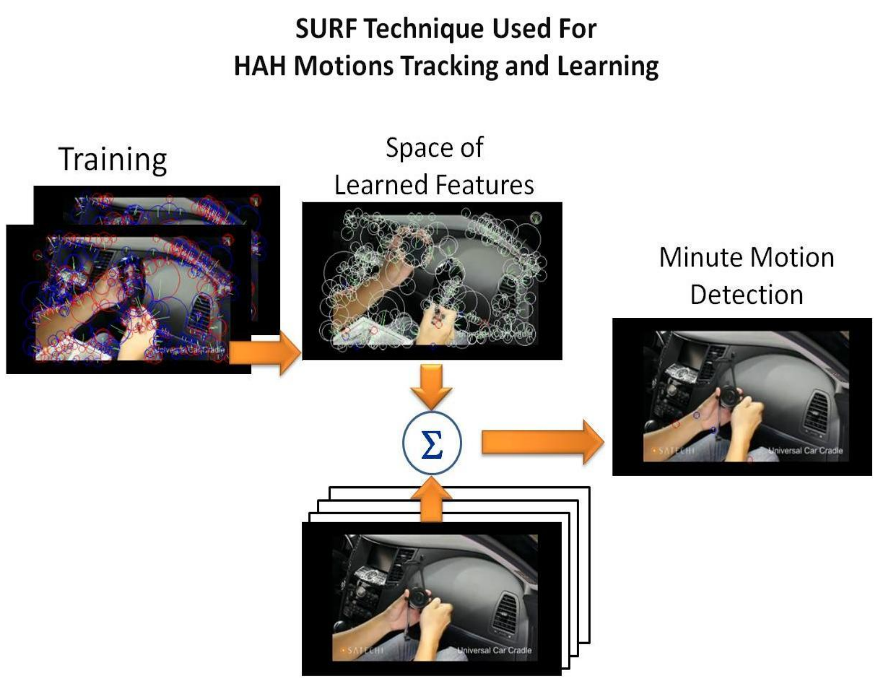

Figure 53. Results for SURF algorithm in IRIS application for detection of foreground objects from background objects.

6.9 Development of Another Robust Algorithm for Hand Color Skin Detection Based on RBG Mean-Shift Technique – To further boost the performance of skin color detection, we developed three other hand skin color detector based on robust cluster technique.

For the purpose, we collected 255 hand gestures in different configuration as illustrated in Figure 54. In addition to these 255 hand pictures, and collected 45 other pictures shows two close or over-lapping hands. The three skin color detectors are based on: Very fast RGB Down Sampling technique, fast RGB Mean-Shift technique, and low RGB Kmeans RGB Clustering technique. Figure 55, 56, and 57 illustrate the resultant of these three techniques in the order aforementioned above. As you may notice, the RGB down sampling technique is fast, yet, the background and foreground are not very separated. The RGB Mean-Shift technique is not as fast as RGB down sampling technique, yet it yielded very respectable results and segmentation of foreground and background is less tedious with this approach. The Kmeans clutter technique is found the slowest of all three techniques, yet it yields very respectable results and foreground and background can be readily separated. The second approach is the best compromised in terms of performance and speed and therefore, we choose this technique for the RGB clustering of hands from the background and isolating them for this configurational characterization via the following algorithms that follow.

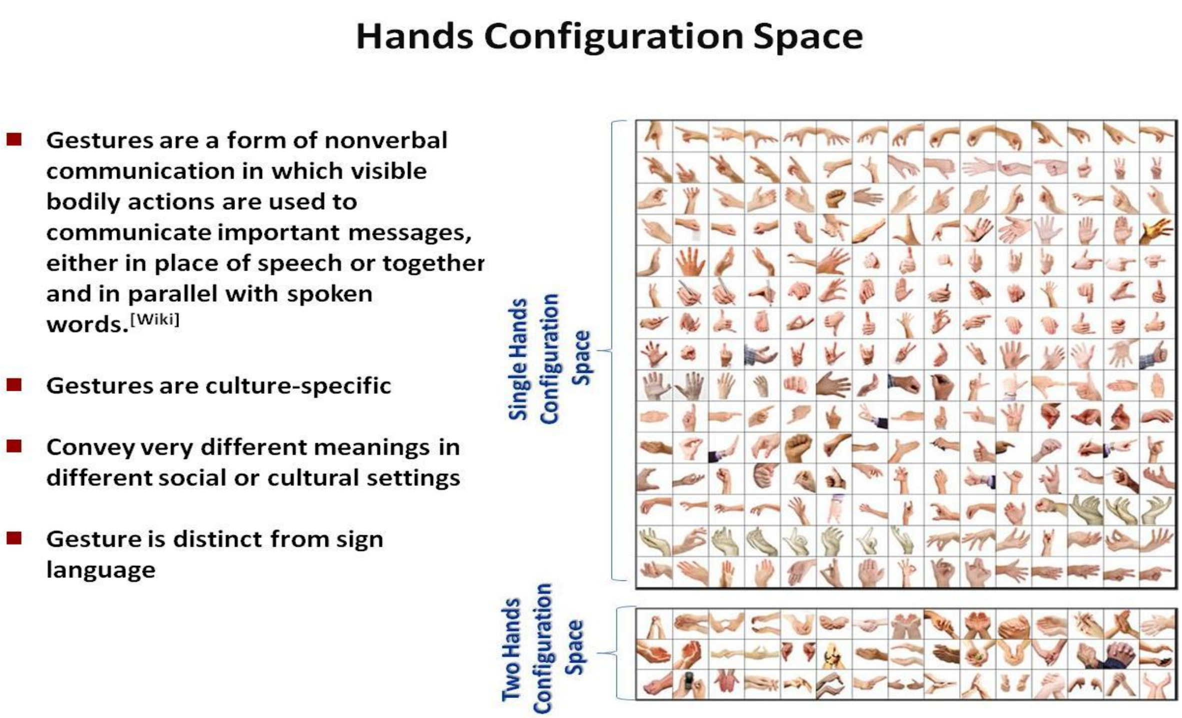

Figure 54. A Sample Space of Selected Hand(s) Configurations for Skin Color Detection and Gesture Training/Learning.

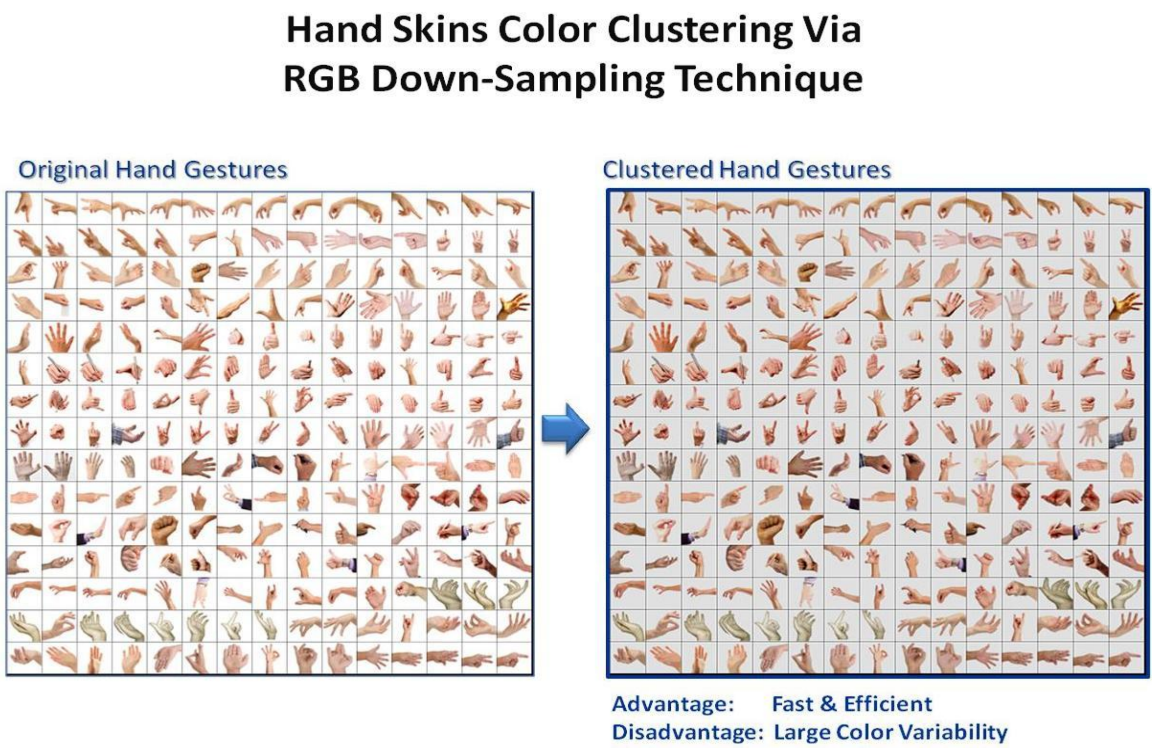

Figure 55. The result of hand training sample space of before (left) and after (right) skin color detection based on very Fast RGB Down Sampling Technique.

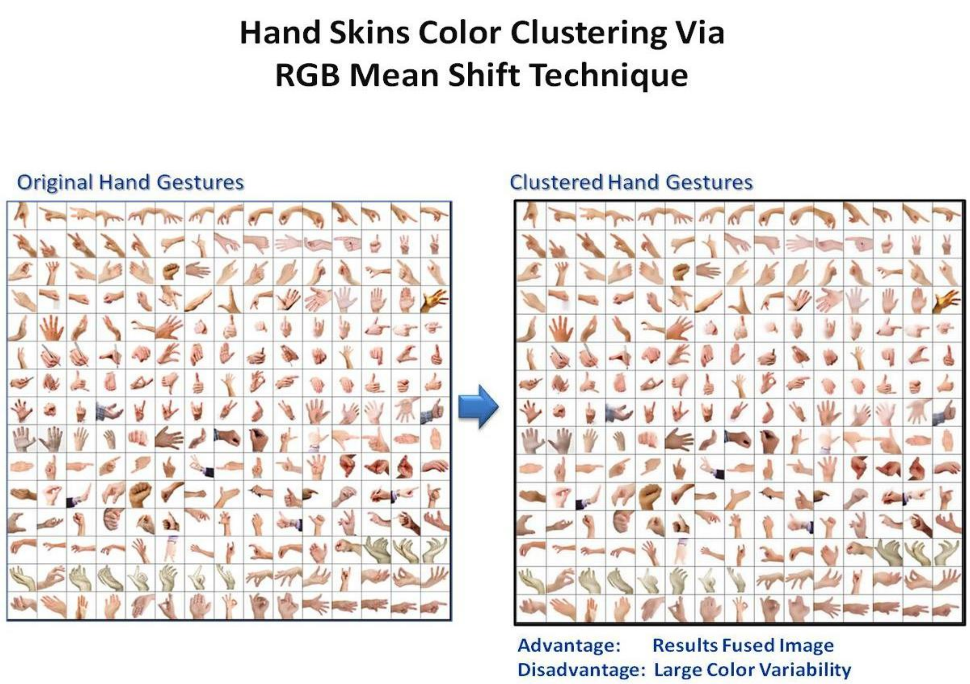

Figure 56. The result of hand training sample space of before (left) and after (right) skin color detection based on Fast RGB Mean Shift Technique.

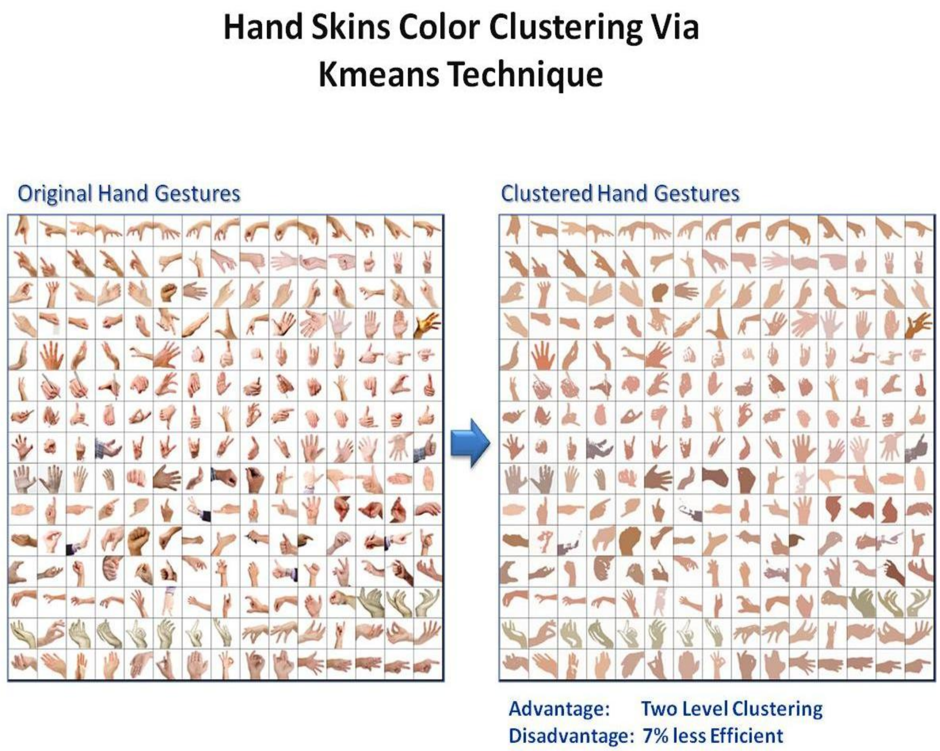

Figure 57. The result of hand training sample space of before (left) and after (right) skin color detection based on Slow Kmeans RGB pixels Cluster.

6.10 Development of Two Competing Techniques for Hand Gesture Classification **–** After detection of hand skin, as demonstrated in the previous sections, we developed two competing techniques for hand gesture classification. The first technique is based on Recursive Hamming Neural Network, and the second technique is based on Random Field Trees. Initially, we labeled the sample space of all our hand training models. We divided our entire hand sample space into twelve different classes. Each class contains a finite number of hand configurations as illustrated in Figures 58, 59, and 60. The twelve hand configuration classes are: Touching, Holding, Writing, Opening, Fisting, Pausing, Gesturing, Cupping, Pinching, Praying, Pointing, and Texting. We note that the class praying contains two-hand formations representing praying hand configuration. We also note that a single hand in any of the given praying configuration is not conclusive of praying. In order to alleviate this misclassification, we resort to consider only two-hand configuration as a state of hand-praying. For the former classifier, namely, hamming neural network (HNN), we generated a batch of invariant features of hand samples for the training of the network. The HNN based on degree of closest of two patterns measured based on hamming distance determine the best class representing a test pattern. Figure 61 presents the Hamming neural network model. In information theory, the Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. In another way, it measures the minimum number of *substitutions* required to change one string into the other, or the minimum number of *errors* that could have transformed one string into the other. The invariant features considered for training of HMM includes seven Hu invariant moments, texture features (variance, energy, entropy, and homogeneity), and three RGB features. For the second classifier, the coefficients of Zernike Polynomial was used as features of hand, and a Random Forest Tree classifier was used for classification of Zernike Polynomial feature vectors. Figure 62 presents the Random Forest Tree Classifier operating based on Zernike Polynomial Coefficients. In mathematics, the Zernike polynomials are a sequence of polynomials that are orthogonal on the unit disk [5]. There are event and odd Zernike polynomials. The even ones are defined as

$$Z_n^m(\rho,\varphi) = R_n^m(\rho)\,\cos(m\,\varphi)$$

and the odd ones as

$$Z_n^{-m}(\rho,\varphi) = R_n^m(\rho)\,\sin(m\,\varphi),$$

where m and n are nonnegative integers with $n \geq m$, φ is the azimuthal angle, ρ is the radial distance $0 \leq \rho \leq 1$, and $R^m{}_n$ are the radial polynomials defined below. Zernike polynomials have the property of being limited to a range of −1 to +1, i.e. $|Z_n^m(\rho,\varphi)| \leq 1$. The radial polynomials $R^m{}_n$ are defined as

$$R_n^m(\rho) = \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k\,(n-k)!}{k!\,\left(\frac{n+m}{2}-k\right)!\,\left(\frac{n-m}{2}-k\right)!}\;\rho^{n-2\,k}$$

for $n - m$ even, and are identically 0 for $n - m$ odd.

Random forests are an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees [6]. In the essence, this random forest technique constructs a collection of decision trees with controlled variance. The training algorithm for random forests applies the general technique of bootstrap aggregating, or bagging, to tree learners. Given a training set $X = x_1, \ldots, x_n$ with responses $Y = y_1$ through y_n, bagging repeatedly selects a bootstrap sample of the training set and fits trees to these samples:

For b = 1 through B:

1. Sample, with replacement, n training examples from X, Y; call these X_b, Y_b.
2. Train a decision or regression tree f_b on X_b, Y_b.

After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x':

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(x')$$

or by taking the majority vote in the case of decision trees.

In the above algorithm, B is a free parameter. Typically, a few hundred to several thousand trees are used, depending on the size and nature of the training set. Increasing the number of trees tends to decrease the variance of the model, without increasing the bias. As a result, the training and test error tend to level off after some number of trees have been fit. An optimal number of trees B can be found using cross-validation, or by observing the out-of-bag error: the mean prediction error on each training sample x_i, using only the trees that did not have x_i in their bootstrap sample. Figure 63 presents the performance evaluation results of these two classifiers. The Random Forest with Zernike Polynomial Coefficient feature vector training has shown 1.9 percent better classification power over the HNN with invariant hand gesture feature vectors.

Figure 58. Four hand configurations: Touching, Holding, Writing, and Opening.

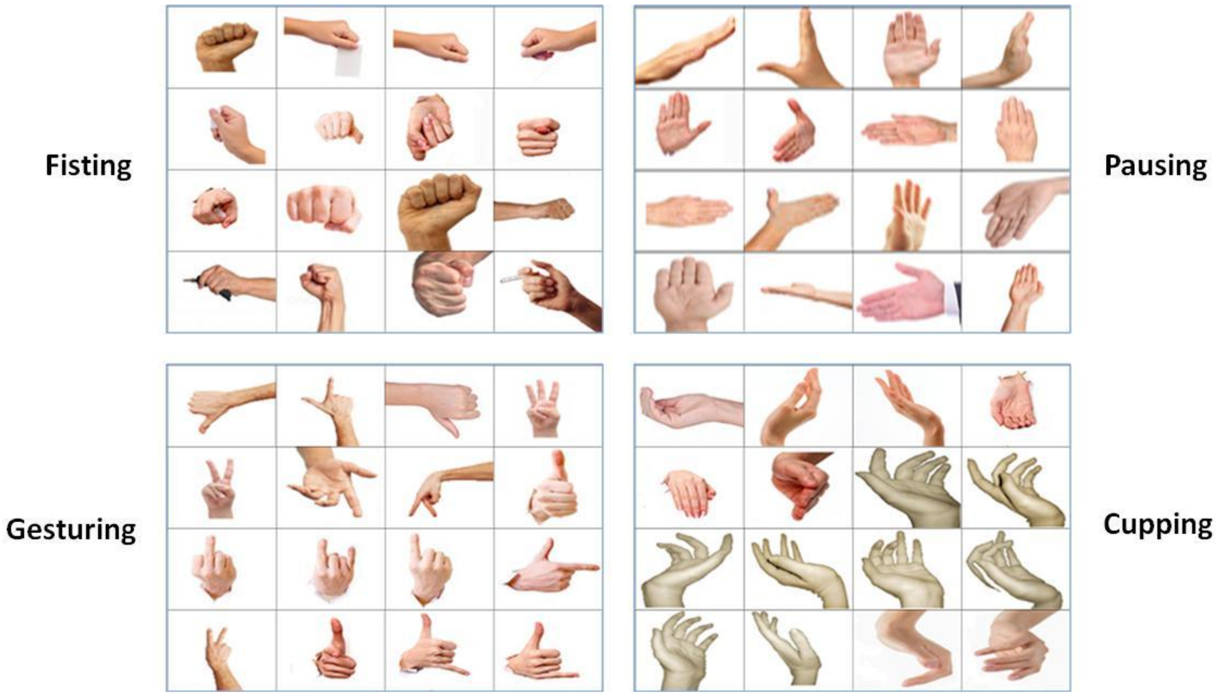

Figure 59. Four hand configurations: Fisting, Pausing, Gesturing, and Cupping.

Figure 60. Four hand configurations: Pinching, Praying, Pointing, and Texting.

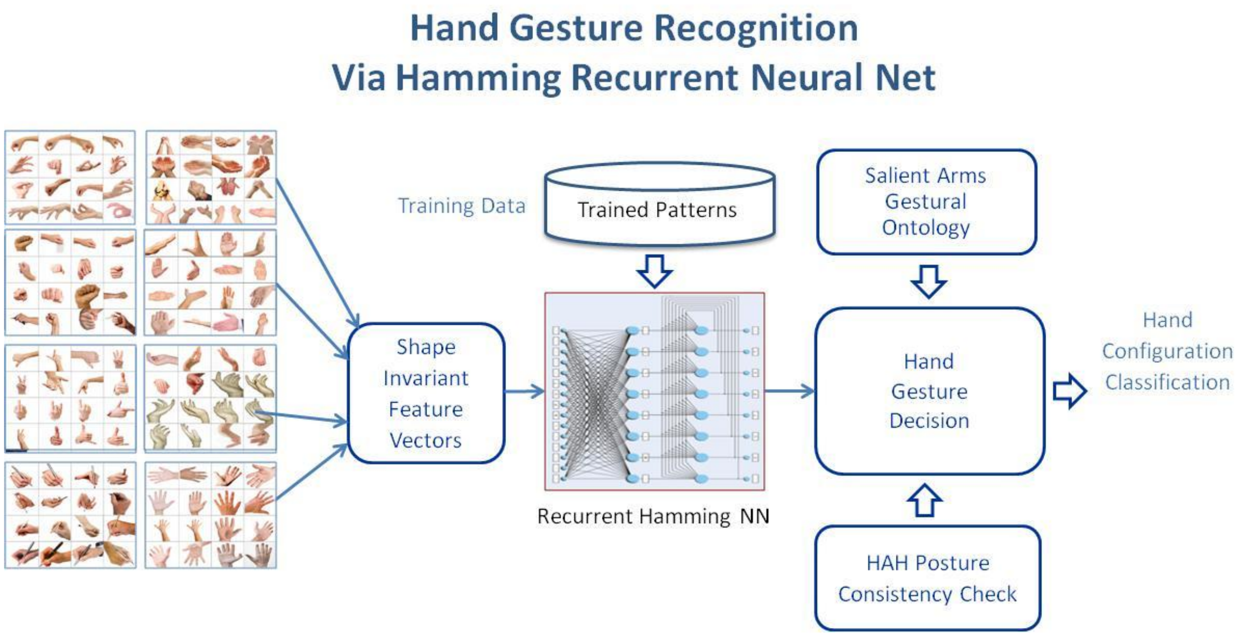

Figure 61. Hamming Neural Network for Classification of Hand Samples.

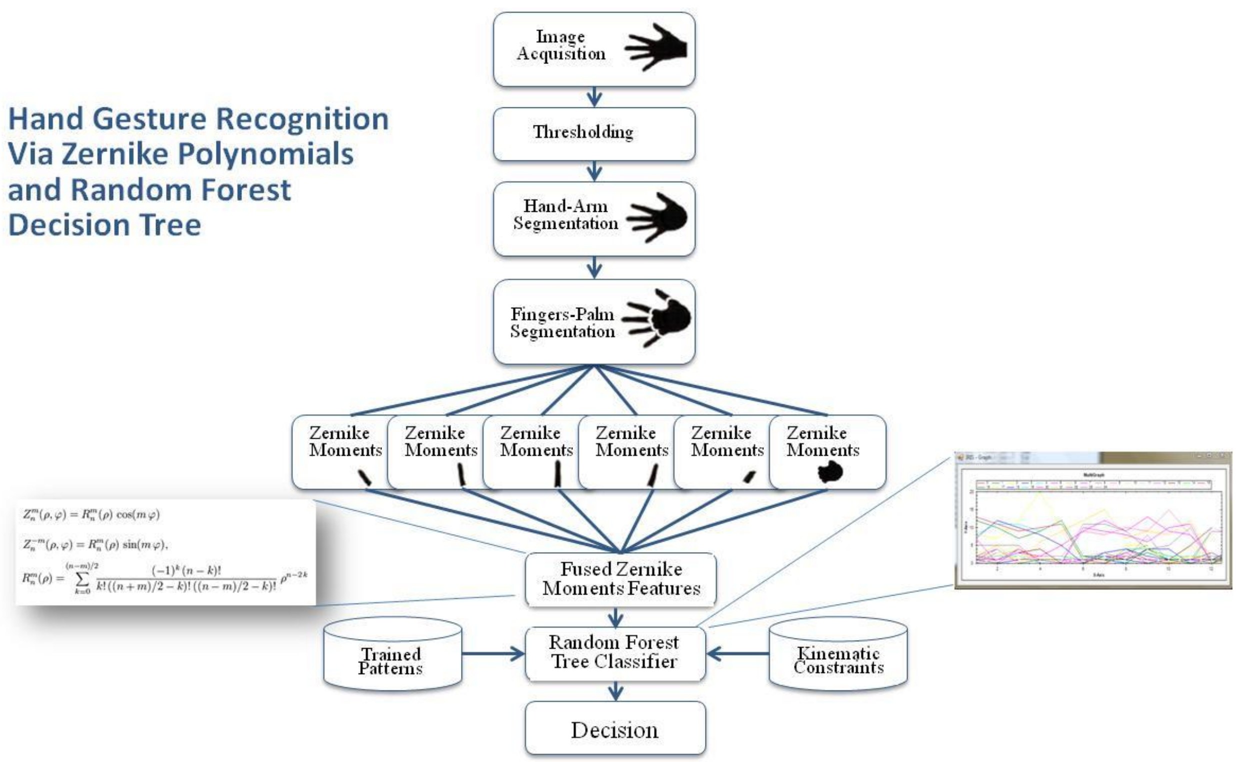

Figure 62. Random Forest Tree Classifier With for Classification of Hand Samples Based on Zernike Polynomials Coefficients.

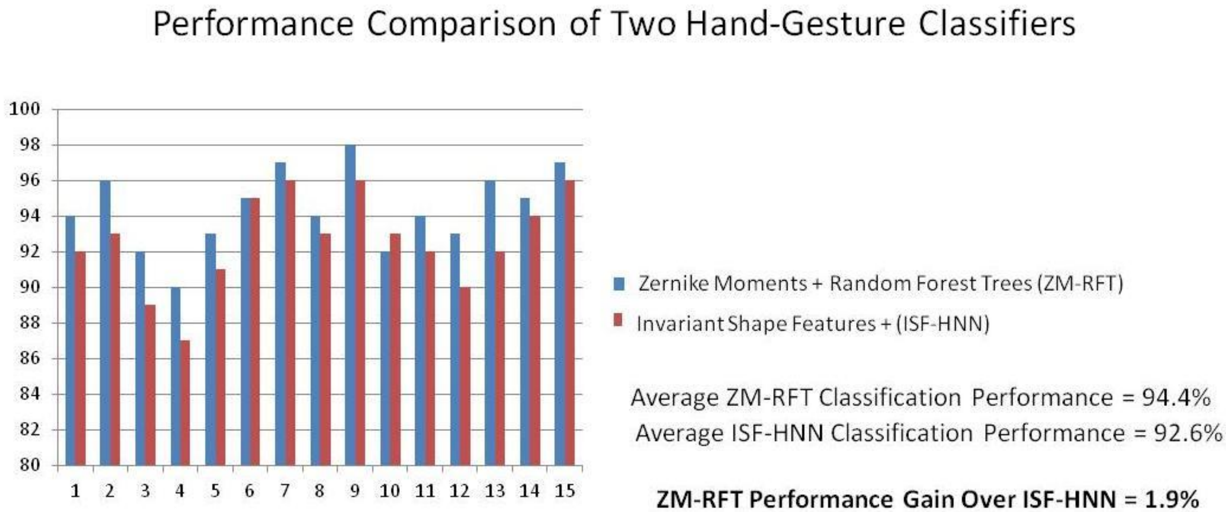

Figure 63. Result of performance comparison of two hand gesture classifiers.

6.11 Development of a Frame Work for Enhancement of This Feasibility Study Toward Phase I of this Project – Based on the learning achieved in this feasibility study project, a new architectural framework was developed as illustrated in Figure 64. This framework considers four main components: Visual Processing, Visual Training, a Gesture Recognition System, and Visual Perceptions Grouping. The Visual processing component detects and tracks the variance of imagery data overtime. The training component, on the other hand, provides the necessary trained taxonomy and ontology for recognition of certain known in-Vehicle activities. The role of Gesture recognition system component is to classify and discriminate normal gestures from anomalous gestures. Each recognized gesture pattern compels a new order of perception that needs to be analyzed and modeled to entail it a salient gesture to annotate. The process begins first, by detecting salient events, then it become a matter of recognizing actions that each action is composed of multiple salient events. An array of associated and correlated actions represents an activity. Therefore, to recognition iVGA, events associated with each occupant needs to be investigated, analyzed, and scrutinized. The proposed framework will be able to fulfill this objective. This architecture allows recommends a fuzzy logic context reasoner as an arbitrator of contextual activities exhibited by the in-vehicle occupants. Furthermore, this architecture framework poses a mechanism for annotating newly discovered information. Humans are more accustomed to reading reporting and understanding situations from written reports. By continual generating *iVGA* annotations, one can trace the results of observed events, actions, and activities overtime and even further interrogate what has happened based on the significance of generated annotations. The annotated reports also offer opportunities for visual data/information analytics and further intelligence processing and pursues.

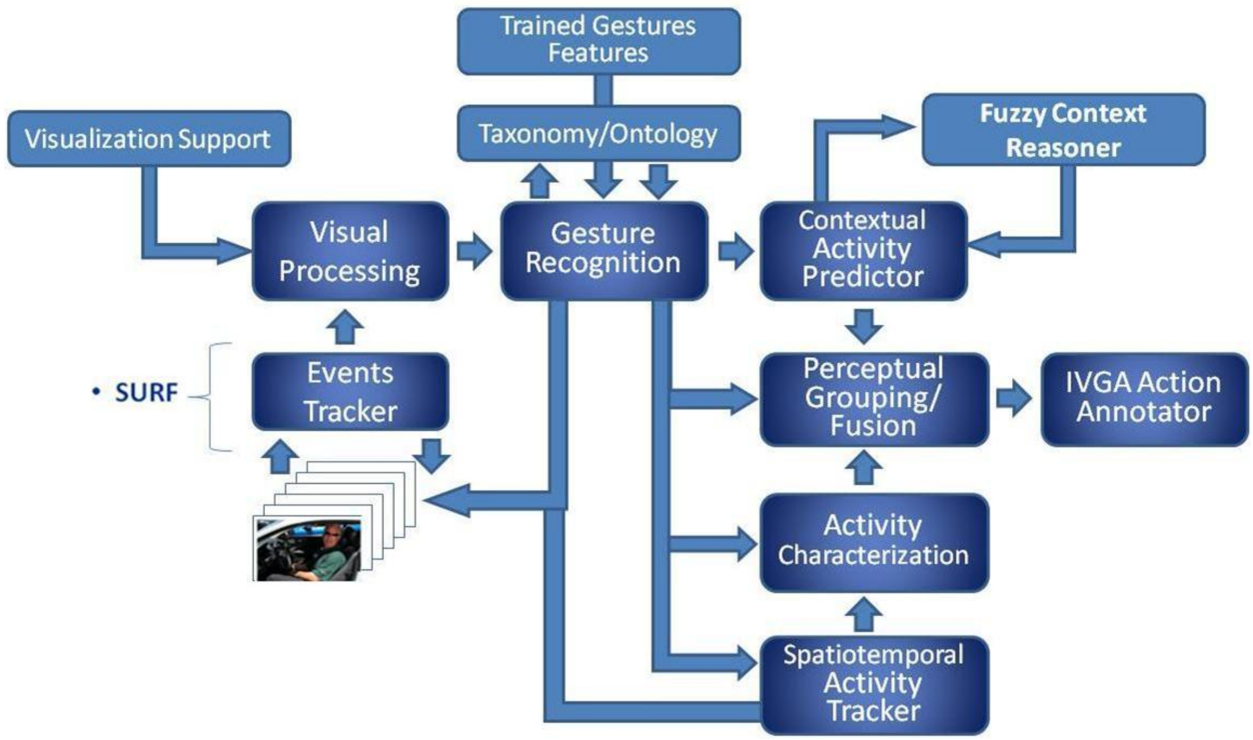

Figure 64. A proposed Architectural framework for in-Vehicle Group Activity (*iVGA*) recognition.

7. Scientific Barriers

There are a number of scientific barriers complicating the spatiotemporal processing of partially observable group activities. These barriers are related to processing of obscured imagery data representing fragmented gestural/postural configurations - specially when they occur in tight confined spaces (e.g., inside a parked vehicle); spatiotemporal understanding of postural gestures (i.e., body, arms, and hands gestures) in order to predict what activities are taking place in confined spaces despite of present of environment clutter; systematic development of apt ontology for discrimination of normal and abnormal behavioral patterns; and spatiotemporal linking of partial information under task uncertainties. Other scientific barriers are related to sustainment of a continuous sensing to perception in lieu of significant operational noise factors and inevitable spatiotemporal variances of gestural and postural configurations.

8. Scientific Significance

The scientific significance of this project is related to adaptive learning, incremental comprehension, constructive perception modeling, and knowledge-based anticipation modeling of group activities occurring in small spaces. This project establishes a sound information-theoretic framework for multi-target characterization, data referencing, opportunistic sensing via robust visual analytics inspired by neurosciences, fuzzy-logic processing, and behavioral patterns learning and modeling, and uncertainty handling via apt information fusion techniques with appropriate performance metrics for evaluation of alternative solutions.

9. Scientific Accomplishments

This project is intended for multiple phases. This report presents our initial feasibility study of this challenging project. In this initial feasibility study, a systematic approach for analysis of target entities operating inside a vehicle, while being observed by a remote surveillance camera, was performed. This analysis established a set of sound ontologies for state-space representations of target entities' postural configurations inside a vehicle as well as their physical limitation – particularly, in terms of kinematic constraints that may result impossible/unrealistic postural configurations of arms during normal activities/operations. To better visualize the admissible and impermissible postural configurations, a virtual model of a humanoid was developed using IRIS software developed by the PI. The humanoid model is a 32-degree-of-freedom robotic model with kinematic motions similar to those of humans. This model was utilized to simulate different human postural configurations in the virtual environment and develop suitable ontology for verification of admissible and inadmissible postural configurations.

Furthermore, we developed suitable image processing (IP) techniques facilitating automatic background segmentation, hands, arms, and head detection and tracking. Several techniques were explored for skin tune detection and segmentation. In occluded spaces, there is large variety of shading variation that compromises the skin tune reflections.

10. Collaboration and Leveraged Funding

The PI is a Co-PI of an ARO-supported MURI project entitled '"Network-based Hard-Soft Sensor Information Fusion". The MURI project is let by University of Buffalo. Dr. John Lavery is the Program Manager of this MURI project. Some funding from MURI is leveraged partially toward support of graduate and undergraduate students currently working on this project. These students are contributing both toward the research objective of project as well as toward the research objective of our on-going MURI project. The MURI project expires in July 31, 2014.

11. Technology Transfer

This project has held 13 monthly teleconference meetings with the ARL technical monitors at ARL. The ARL technical monitors on this project are: Dr. Alex Chan, and Dr. Shuowen Hu. This frequent interaction with ARL has been significantly instrumental to transfer technology and this cooperative technology transfer effort will be maintained to assure the successful achievement of this mission.

12. Future Research Plan

In the second phase of this project we plan to develop a fuzzy-logic-based inference engine for inferring postural configurations based on spatiotemporally processed visual imagery data. This fuzzy inference engine enables fuzzy reasoning of postural configurations under visual perception uncertainties. In the third phase of this project we will develop the proposed multi-layer Hidden Markov Model, for probabilistic state-transition, and hence group activity recognition under partial visual observation constraints. At final stage of this project, we will develop a system for semantic annotation of group activities occurring in confined spaces.

13. Anticipated Scientific Accomplishments

The anticipated scientific accomplishments include: (1) robust image processing, dynamic body (i.e., head, arms, and hands) postural/gestural configuration tracking under occlusion and imagery constraints, (2) fuzzy reasoning of spatiotemporal physical postures/gestures for normal and anomaly discriminatory behavioral pattern recognition, (3) dynamic group activity recognition via probabilistic multi-layer Hidden Markov Model, and (4) semantic annotation of group activities in confined spaces.

14. Students Involvements

This project has involved a number of high-school, undergraduate, Master, and Ph.D. students including

- Vinayak Elangovan (Ph.D.) - Part time Spring 2014
- Azin Poshtkar, (Master Grad) Full time Spring and Summer 2014
- Ayele Tegegne (UG) - Part time, in Summer 2014, a Navy Veteran
- David A. Potter Jr. (UG) - Part time, in Summer 2014 – An Army Veteran
- Pedro Henrique Tavares (UG) - Part time, in Summer 2014
- Brent Warner (UG) – Part time, in Summer 2014, An Army Helicopter Veteran
- Ramon Gonzalez (UG) - Part time, in Summer 2013, Graduated in spring 2014
- Daniel Allen (UG) - Part time, in Summer 2013, Graduated in spring 2014

High School Summer Interns

- Jalyn Edmundson (Lighthouse Christian School High-School) , Part Time in Summer 2013
- Freeman Johnson (Martin Luther King Magnet High-School) , Part Time in Summer 2013
- Zheer Ahmed (Martin Luther King Magnet High-School) , Part Time in Summer 2013

15. Conclusion

This project presents a technical approach for detection and recognition of *in-Vehicle Group Activities* (*iVGA*) in confined obstructed spaces. Robustness in recognition of such activities can significantly benefit home-land security as well as battlefield automated surveillance and reconnaissance operations. Particularly, this project establishes a technical bridge for achievement of this goal.

16. References

[1] Quinlan, J. R. 1986. Induction of Decision Trees. Mach. Learn. 1, 1 (Mar. 1986), 81-106.

[2] Mitchell, Tom M. *Machine Learning*. McGraw-Hill, 1997. pp. 55–58.

[3] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346–359, 2008.

[4] Finkelstein, Amir; Neeratsharma, "Double Integrals By Summing Values Of Cumulative Distribution Function". *Wolfram Demonstration Project*, 2010.

[5] Díaz, J. A.; Fernández-Dorado, J.; Pizarro, C.; Arasa, J., "Zernike Coefficients for Concentric, Circular, Scaled Pupils: An Equivalent Expression". *Journal of Modern Optics* 56 (1): 149–155, 2009.

[6] Breiman, Leo, "Random Forests", Machine Learning *Vol.* 45 (1): 5–32, 2001.

[7] Habibi, M. S., and Shirkhodaie, A., "Multi-attributed tagged big data exploitation for hidden concepts discovery", SPIE Defense, Security and Sensing Conference, Baltimore, Maryland, April (2014).

[8] Elangovan, V., and Shirkhodaie, A., "Knowledge discovery in group activities through sequential observation analysis", SPIE Defense, Security and Sensing Conference, Baltimore, Maryland, April (2014).

[9] Alkilani, A., and Shirkhodaie, A., "Acoustic events semantic detection, classification, and annotation for persistent surveillance applications", SPIE Defense, Security and Sensing Conference, Baltimore, Maryland, April (2014).

[10] Shirkhodaie, A., Elangovan, V., Habibi, M. S., and Alkilani, A., "A Decision Support System for Fusion of Soft and Hard-sensor Information Based on Latent Semantic Analysis Technique", SPIE Defense, Security and Sensing Conference, Baltimore, Maryland, April (2013).

[11] Elangovan, V., and Shirkhodaie, A., "A Robust Technique for Group Activities Recognition Based on Fusion of Extracted Features in Video Streams", SPIE Defense, Security and Sensing Conference, Baltimore, Maryland, April (2013).

[12] Elangovan, V., Bashir, A., and Shirkhodaie, A., "A Multi-attribute Based Methodology for Vehicle Detection & Identification", SPIE Defense, Security and Sensing Conference, Baltimore, Maryland, April (2013).

[13] Elangovan, V., Alkilani, A., and Shirkhodaie, A., "A Multi-Modality Attributes Unmasking Scheme for Group Activity Characterization and Data Fusion", IEEE Intelligence and Security Informatics (ISI), Seattle, WA, June (2013).

[14] Alkilani, A., and Shirkhodaie, A., "Acoustic Recognition of Human-Object Interactions in Persistent Surveillance Systems", SPIE Defense, Security and Sensing Conference, Baltimore, Maryland, April (2013).

[15] Habibi, M. S., and Shirkhodaie, A., "Mining Patterns in Persistent Surveillance Systems with Smart Query and Visual Analytics", SPIE Defense, Security and Sensing Conference, Baltimore, Maryland, April (2013).

[16] Elangovan V., and Shirkhodaie, A., "Recognition of Human Activity Characteristics based on

State Transitions Modeling Technique," SPIE Defense, Security and Sensing, Baltimore, Maryland, April (2012).

[17] Elangovan V., Bandaru V.K., and Shirkhodaie, A., "Team Activity Analysis and Recognition Based on Kinect Depth Map and Optical Imagery Techniques," SPIE Defense, Security and Sensing, Baltimore, Maryland, April (2012).

[18] Alkilani, A., and Shirkhodaie, A.,"A survey on acoustic signature recognition and classification techniques for persistent surveillance systems", SPIE 2012 Defense, Security, and Sensing Symposium, Conference:Signal Processing, Sensor Fusion, and Target Recognition XXI, Paper number 8392-28, Baltimore, Maryland, April (2012).

[19] Shirkhodaie, A., "Perceptual Semantic Labeling of Human-Vehicle Interactions (HVI)," 2nd Annual Human, Light Vehicle and Tunnel Detection Workshop, May 3-4, 2012, Baltimore, MD.

[20] Shirkhodaie, A., "Semantic Labeling of Human-Vehicle Interactions Via Acoustic Events Characterization and Inference," 2nd Annual Human, Light Vehicle and Tunnel Detection Workshop, May 3-4, 2012, Baltimore, MD.